

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/031588 A2

(51) International Patent Classification⁷:

C12N

(21) International Application Number: PCT/US02/32512

(22) International Filing Date: 10 October 2002 (10.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/328,655 11 October 2001 (11.10.2001) US
60/363,774 13 March 2002 (13.03.2002) US

(71) Applicants (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. [IT/IT]; VIA PONTINA KM. 30.600, I-00040 POMEZIA (IT).

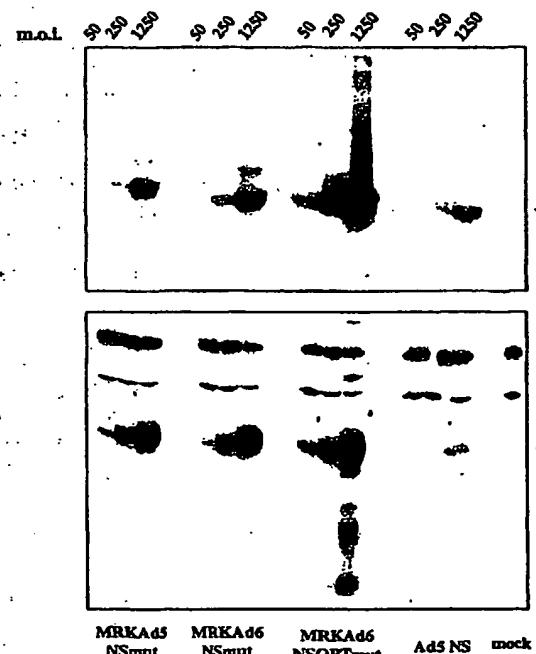
(72) Inventors; and

(75) Inventors/Applicants (for US only): EMINI, Emilio, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). KASLOW, David, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). BETT, Andrew, J. [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). NICOSIA, Alfredo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LAHM, Armin [DE/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LUZZAGO, Alessandra [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). CORTESE, Riccardo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). COLLOCA, Stefano [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT).

(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

[Continued on next page]

(54) Title: HEPATITIS C VIRUS VACCINE



(57) Abstract: The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (m.o.i.; indicated at the top) with Adenovectors expressing the different HCV NS casettes. Mature NS5B and NS5A products were detected with specific antibodies.

WO 03/031588 A2

BEST AVAILABLE COPY



(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM).

European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION
HEPATITIS C VIRUS VACCINE

RELATED APPLICATIONS

5 The present application claims priority to provisional applications U.S. Serial No. 60/363,774, filed March 13, 2002, and U.S. Serial No. 60/328,655, filed October 11, 2001, each of which are hereby incorporated by reference herein.

BACKGROUND OF THE INVENTION

10 The references cited in the present application are not admitted to be prior art to the claimed invention.

About 3% of the world's population are infected with the Hepatitis C virus (HCV). (Wasley *et al.*, *Semin. Liver Dis.* 20, 1-16, 2000.) Exposure to HCV results in an overt acute disease in a small percentage of cases, while in most 15 instances the virus establishes a chronic infection causing liver inflammation and slowly progresses into liver failure and cirrhosis. (Iwarson, *FEMS Microbiol. Rev.* 14, 201-204, 1994.) In addition, epidemiological surveys indicate an important role of HCV in the pathogenesis of hepatocellular carcinoma. (Kew, *FEMS Microbiol. Rev.* 14, 211-220, 1994, Alter, *Blood* 85, 1681-1695, 1995.)

20 Prior to the implementation of routine blood screening for HCV in 1992, most infections were contracted by inadvertent exposure to contaminated blood, blood products or transplanted organs. In those areas where blood screening of HCV is carried out, HCV is primarily contracted through direct percutaneous exposure to infected blood, *i.e.*, intravenous drug use. Less frequent methods of transmission 25 include perinatal exposure, hemodialysis, and sexual contact with an HCV infected person. (Alter *et al.*, *N. Engl. J. Med.* 341(8), 556-562, 1999, Alter, *J. Hepatol.* 31 *Suppl.* 88-91, 1999. *Semin. Liver. Dis.* 201, 1-16, 2000.)

The HCV genome consists of a single strand RNA about 9.5 kb encoding a precursor polyprotein of about 3000 amino acids. (Choo *et al.*, *Science* 244, 362-364, 1989, Choo *et al.*, *Science* 244, 359-362, 1989, Takamizawa *et al.*, *J. Virol.* 65, 1105-1113, 1991.) The HCV polyprotein contains the viral proteins in the order: C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B.

Individual viral proteins are produced by proteolysis of the HCV polyprotein. Host cell proteases release the putative structural proteins C, E1, E2, and

p7, and create the N-terminus of NS2 at amino acid 810. (Mizushima *et al.*, *J. Virol.* 68, 2731-2734, 1994, Hijikata *et al.*, *P.N.A.S. USA* 90, 10773-10777, 1993.)

The non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B presumably form the virus replication machinery and are released from the polyprotein. A zinc-dependent protease associated with NS2 and the N-terminus of NS3 is responsible for cleavage between NS2 and NS3. (Grakoui *et al.*, *J. Virol.* 67, 1385-1395, 1993, Hijikata *et al.*, *P.N.A.S. USA* 90, 10773-10777, 1993.) A distinct serine protease located in the N-terminal domain of NS3 is responsible for proteolytic cleavages at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B junctions. (Bartenschlager *et al.*, *J. Virol.* 67, 3835-3844, 1993, Grakoui *et al.*, *Proc. Natl. Acad. Sci. USA* 90, 10583-10587, 1993, Tomei *et al.*, *J. Virol.* 67, 4017-4026, 1993.) NS4A provides a cofactor for NS3 activity. (Failla *et al.*, *J. Virol.* 68, 3753-3760, 1994, De Francesco *et al.*, U.S. Patent No. 5,739,002.)

NS5A is a highly phosphorylated protein conferring interferon resistance. (De Francesco *et al.*, *Semin. Liver Dis.*, 20(1), 69-83, 2000, Pawlotsky, *Viral Hepat. Suppl. 1*, 47-48, 1999.)

NS5B provides an RNA-dependent RNA polymerase. (De Francesco *et al.*, International Publication Number WO 96/37619, Behrens *et al.*, *EMBO* 15, 12-22, 1996, Lohmann *et al.*, *Virology* 249, 108-118, 1998.)

20

SUMMARY OF THE INVENTION

The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

A HCV specific CMI response refers to the production of cytotoxic T lymphocytes and T helper cells that recognize an HCV antigen. The CMI response may also include non-HCV specific immune effects.

Preferred nucleic acids encode a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that is substantially similar to SEQ. ID. NO. 1 and has sufficient protease activity to process itself to produce at least a polypeptide substantially similar to the NS5B region present in SEQ. ID. NO. 1. The produced polypeptide corresponding to NS5B is enzymatically inactive. More preferably, the HCV polypeptide has sufficient

protease activity to produce polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

Reference to a "substantially similar sequence" indicates an identity of at least about 65% to a reference sequence. Thus, for example, polypeptides having an amino acid sequence substantially similar to SEQ. ID. NO. 1 have an overall amino acid identity of at least about 65% to SEQ. ID. NO. 1.

Polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B have an amino acid sequence identity of at least about 65% to the corresponding region in SEQ. ID. NO. 1. Such corresponding polypeptides are also referred to 10 herein as NS3, NS4A, NS4B, NS5A, and NS5B polypeptides.

Thus, a first aspect of the present invention describes a nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The encoded polypeptide has sufficient protease activity to process itself to produce an NS5B polypeptide that is 15 enzymatically inactive.

In a preferred embodiment, the nucleic acid is an expression vector capable of expressing the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide in a desired human cell. Expression inside a human cell has therapeutic applications for 20 actively treating an HCV infection and for prophylactically treating against an HCV infection.

An expression vector contains a nucleotide sequence encoding a polypeptide along with regulatory elements for proper transcription and processing. The regulatory elements that may be present include those naturally associated with the nucleotide sequence encoding the polypeptide and exogenous regulatory elements 25 not naturally associated with the nucleotide sequence. Exogenous regulatory elements such as an exogenous promoter can be useful for expression in a particular host, such as in a human cell. Examples of regulatory elements useful for functional expression include a promoter, a terminator, a ribosome binding site, and a polyadenylation signal.

30 Another aspect of the present invention describes a nucleic acid comprising a gene expression cassette able to express in a human cell a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The polypeptide can process itself to produce an enzymatically inactive NS5B protein. The gene expression cassette contains at least the following:

5 a) a promoter transcriptionally coupled to a nucleotide sequence
encoding a polypeptide;
b) a 5' ribosome binding site functionally coupled to the nucleotide
sequence,
c) a terminator joined to the 3' end of the nucleotide sequence, and
d) a 3' polyadenylation signal functionally coupled to the nucleotide
sequence.

Reference to "transcriptionally coupled" indicates that the promoter is positioned such that transcription of the nucleotide sequence can be brought about by RNA polymerase binding at the promoter. Transcriptionally coupled does not require that the sequence being transcribed is adjacent to the promoter.

Reference to "functionally coupled" indicates the ability to mediate an effect on the nucleotide sequence. Functionally coupled does not require that the coupled sequences be adjacent to each other. A 3' polyadenylation signal functionally coupled to the nucleotide sequence facilitates cleavage and polyadenylation of the transcribed RNA. A 5' ribosome binding site functionally coupled to the nucleotide sequence facilitates ribosome binding.

In preferred embodiments the nucleic acid is a DNA plasmid vector or an adenovector suitable for either therapeutic application in treating HCV or as an intermediate in the production of a therapeutic vector. Treating HCV includes actively treating an HCV infection and prophylactically treating against an HCV infection.

25 Another aspect of the present invention describes an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1 that is produced by a process involving (a) homologous recombination and (b) adenovector rescue. The homologous recombinant step produces an adenovirus genome plasmid. The adenovector rescue step produces the adenovector from the adenogenome plasmid.

30 Adenovirus genome plasmids described herein contain a recombinant adenovirus genome having a deletion in the E1 region and optionally in the E3 region and a gene expression cassette inserted into one of the deleted regions. The recombinant adenovirus genome is made of regions substantially similar to one or more adenovirus serotypes.

Another aspect of the present invention describes an adenovector
35 consisting of the nucleic acid sequence of SEQ. ID. NO. 4 or a derivative thereof,

wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ. ID. NO. 4 replaced with the HCV polyprotein encoding sequence of either SEQ. ID. NO. 3, SEQ. ID. NO. 10 or SEQ. ID. NO. 11.

Another aspect of the present invention describes a cultured recombinant cell comprising a nucleic acid containing a sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The recombinant cell has a variety of uses such as being used to replicate nucleic acid encoding the polypeptide in vector construction methods.

Another aspect of the present invention describes a method of making an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1. The method involves the steps of (a) producing an adenovirus genome plasmid containing a recombinant adenovirus genome with deletions in the E1 and E3 regions and a gene expression cassette inserted into one of the deleted regions and (b) rescuing the adenovector from the adenovirus genome plasmid.

Another aspect of the present invention describes a pharmaceutical composition comprising a vector for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1 and a pharmaceutically acceptable carrier. The vector is suitable for administration and polypeptide expression in a patient.

A "patient" refers to a mammal capable of being infected with HCV. A patient may or may not be infected with HCV. Examples of patients are humans and chimpanzees.

Another aspect of the present invention describes a method of treating a patient comprising the step of administering to the patient an effective amount of a vector expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The vector is suitable for administration and polypeptide expression in the patient.

The patient undergoing treatment may or may not be infected with HCV. For a patient infected with HCV, an effective amount is sufficient to achieve one or more of the following effects: reduce the ability of HCV to replicate, reduce HCV load, increase viral clearance, and increase one or more HCV specific CMI responses. For a patient not infected with HCV, an effective amount is sufficient to achieve one or more of the following: an increased ability to produce one or more components of a HCV specific CMI response to a HCV infection, a reduced

susceptibility to HCV infection, and a reduced ability of the infecting virus to establish persistent infection for chronic disease.

Another aspect of the present invention features a recombinant nucleic acid comprising an Ad6 region and a region not present in Ad6. Reference to 5 "recombinant" nucleic acid indicates the presence of two or more nucleic acid regions not naturally associated with each other. Preferably, the Ad6 recombinant nucleic acid contains Ad6 regions and a gene expression cassette coding for a polypeptide heterologous to Ad6.

10 Other features and advantages of the present invention are apparent from the additional descriptions provided herein including the different examples. The provided examples illustrate different components and methodology useful in practicing the present invention. The examples do not limit the claimed invention. Based on the present disclosure the skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

15

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B illustrate SEQ. ID. NO. 1.

20 Figures 2A, 2B, 2C, and 2D illustrate SEQ. ID. NO. 2. SEQ. ID. NO. 2 provides a nucleotide sequence coding for SEQ. ID. NO. 1 along with an optimized internal ribosome entry site and TAAA termination. Nucleotides 1-6 provides an optimized internal ribosome entry site. Nucleotides 7-5961 code for a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with nucleotides in positions 5137 to 5145 providing a AlaAlaGly sequence in amino acid positions 1711 to 1713 that renders NS5B inactive. Nucleotides 5962-5965 provide a TAAA termination.

25

Figures 3A, 3B, 3C, and 3D illustrate SEQ. ID. NO. 3. SEQ. ID. NO. 3 is a codon optimized version of SEQ. ID. NO. 2. Nucleotides 7-5961 encode a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

30

Figures 4A-4M illustrate MRKAd6-NSmut (SEQ. ID. NO. 4). SEQ. ID. NO. 4 is an adenovector containing an expression cassette where the polypeptide of SEQ. ID. NO. 1 is encoded by SEQ. ID. NO. 2. Base pairs 1-450 correspond to the Ad5 bp 1 to 450; base pairs 462 to 1252 correspond to the human CMV promoter; base pairs 1258 to 1267 correspond to the Kozak sequence; base pairs 1264 to 7222 correspond to the NS genes; base pairs 7231 to 7451 correspond to the BGH polyadenylation signal; base pairs 7469 to 9506 correspond to Ad5 base pairs 3511 to 5548; base pairs 9507 to 32121 correspond to Ad6 base pairs 5542 to 28156; base

pairs 32122 to 35117 correspond to Ad6 base pairs 30789 to 33784; and base pairs 35118 to 37089 correspond to Ad5 base pairs 33967 to 35935.

Figures 5A-5O illustrate SEQ. ID. NOs. 5 and 6. SEQ. ID. NO. 5 encodes a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with an active 5 RNA dependent RNA polymerase. SEQ. ID. NO. 6 provides the amino acid sequence for the polypeptide.

Figures 6A-6C provide the nucleic acid sequence for pV1JnsA (SEQ. ID. NO. 7).

Figures 7A-7N provide the nucleic acid sequence for the Ad6 genome 10 (SEQ. ID. NO. 8).

Figures 8A-8K provide the nucleic acid sequence for the Ad5 genome (SEQ. ID. NO. 9).

Figure 9 illustrates different regions of the Ad6 genome. The linear (35759 bp) ds DNA genome is indicated by two parallel lines and is divided into 100 15 map units. Transcription units are shown relative to their position and orientation in the genome. Early genes (E1A, E1B, E2A/B, E3 and E4 are indicated by gray arrows. Late genes (L1 to L5), indicated by black arrows, are produced by alternative splicing of a transcript produced from the major late promoter (MLP) and all contain the tripartite leader (1, 2, 3) at their 5' ends. The E1 region is located from approximately 20 1.0 to 11.5 map units, the E2 region from 75.0 to 11.5 map units, E3 from 76.1 to 86.7 map units, and E4 from 99.5 to 91.2 map units. The major late transcription unit is located between 16.0 and 91.2 map units.

Figure 10 illustrates homologous recombination to recover pAdE1-E3+ containing Ad6 and Ad5 regions.

Figure 11 illustrates homologous recombinant to recover a pAdE1-E3+ containing Ad6 regions.

Figure 12 illustrates a western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing different HCV NS cassettes. Mature NS3 and NS5A products were detected with specific antibodies. "pV1Jns-NS" refers 30 to a pV1JnsA plasmid where a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is encoded by SEQ. ID. NO. 5, and SEQ. ID. NO. 5 is inserted between bases 1881 and 1912 of SEQ. ID. NO. 7. "pV1Jns-NSmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 2 is inserted between bases 1882 and 1925 of SEQ. ID. NO. 7. "pV1Jns-NSOPTmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 3 is inserted between 35 bases 1881 and 1905 of SEQ. ID. NO. 7.

Figures 13A and 13B illustrate T cell responses by IFN γ ELIspot induced in C57black6 mice (A) and BalbC mice (B) by two injections of 25 μ g and 50 μ g, respectively, of plasmid DNA encoding the different HCV NS cassettes with Gene Electro-Transfer (GET).

5 Figure 14 illustrates protein expression from different adenovectors upon infection of HeLa cells. MRKAd5-NSmut is an adenovector based on an Ad5 sequence (SEQ. ID. NO. 9), where the Ad5 genome has an E1 deletion of base pairs 451 to 3510, an E3 deletion of base pairs 28134 to 30817, and has the NS3-NS4A-NS4B-NS5A-NS5B expression cassette as provided in base pairs 451 to 7468 of SEQ. 10 ID. NO. 4 inserted between positions 450 and 3511. Ad5-NS is an adenovector based on an Ad5 backbone with an E1 deletion of base pairs 342 to 3523, and E3 deletion of base pairs 28134 to 30817 and containing an expression cassette encoding a NS3-NS4A-NS4B-NS5A-NS5B from SEQ. ID. NO. 5. "MRKAd6-NSOPTmut" refers to an adenovector having a modified SEQ. ID. NO. 4 sequence, wherein base pairs 1258 15 to 7222 of SEQ. ID. NO. 4 is replaced with SEQ. ID. NO. 3.

Figure 15 illustrates T cell responses by IFN γ ELIspot induced in C57black6 mice by two injections of 10⁹ vp of adenovectors containing different HCV non-structural gene cassettes.

20 Figures 16A-16D illustrate T cell responses by IFN γ ELIspot induced in Rhesus monkeys by one or two injections of 10¹⁰ vp (A) or 10¹¹ vp (B) of adenovectors containing different HCV non-structural gene cassettes.

Figures 17A and 17B illustrates CD8+ T cell responses by IFN γ ICS induced in Rhesus monkeys by two injections of 10¹⁰ vp (A) or 10¹¹ vp (B) of adenovectors encoding the different HCV non-structural gene cassettes.

25 Figures 18A-18F illustrate T cell responses by bulk CTL assay induced in Rhesus monkeys by two injections of 10¹¹ vp of Ad5-NS (A), MRKAd5-NSmut (B), or MRKAd6-NSmut (C).

Figure 19 illustrates the plasmid pE2.

30 Figures 20A-D illustrates the partial codon optimized sequence NSsuboptmut (SEQ. ID. NO. 10). Coding sequence for the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is from base 7 to 5961.

DETAILED DESCRIPTION OF THE INVENTION

The present invention features Ad6 vectors and nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that contains an inactive NS5B region. Providing an inactive NS5B region supplies NS5B antigens while reducing the possibility of adverse side effects due to an active viral RNA polymerase. Uses of the featured nucleic acid include use as a vaccine component to introduce into a cell an HCV polypeptide that provides a broad range of antigens for generating a CMI response against HCV, and as an intermediate for producing such a vaccine component.

The adaptive cellular immune response can function to recognize viral antigens in HCV infected cells throughout the body due to the ubiquitous distribution of major histocompatibility complex (MHC) class I and II expression, to induce immunological memory, and to maintain immunological memory. These functions are attributed to antigen-specific CD4+ T helper (Th) and CD8+ cytotoxic T cells (CTL).

Upon activation via their specific T cell receptors, HCV specific Th cells fulfill a variety of immunoregulatory functions, most of them mediated by Th1 and Th2 cytokines. HCV specific Th cells assist in the activation and differentiation of B cells and induction and stimulation of virus-specific cytotoxic T cells. Together with CTL, Th cells may also secrete IFN- γ and TNF- α that inhibit replication and gene expression of several viruses. Additionally, Th cells and CTL, the main effector cells, can induce apoptosis and lysis of virus infected cells.

HCV specific CTL are generated from antigens processed by professional antigen presenting cells (pAPCs). Antigens can be either synthesized within or introduced into pAPCs. Antigen synthesis in a pAPC can be brought about by introducing into the cell an expression cassette encoding the antigen.

A preferred route of nucleic acid vaccine administration is an intramuscular route. Intramuscular administration appears to result in the introduction and expression of nucleic acid into somatic cells and pAPCs. HCV antigens produced in the somatic cells can be transferred to pAPCs for presentation in the context of MHC class I molecules. (Donnelly *et al.*, *Annu. Rev. Immunol.* 15:617-648, 1997.)

pAPCs process longer length antigens into smaller peptide antigens in the proteasome complex. The antigen is translocated into the endoplasmic reticulum/Golgi complex secretory pathway for association with MHC class I

proteins. CD8+ T lymphocytes recognize antigen associated with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein.

Using a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide as a vaccine component allows for production of a broad range of 5 antigens capable of generating CMI responses from a single vector. The polypeptide should be able to process itself sufficiently to produce at least a region corresponding to NS5B. Preferred nucleic acids encode an amino acid sequence substantially similar to SEQ. ID. NO. 1 that has sufficient protease activity to process itself to produce individual HCV polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, 10 and NS5B regions present in SEQ. ID. NO. 1.

A polypeptide substantially similar to SEQ. ID. NO. 1 with sufficient protease activity to process itself in a cell provides the cell with T cell epitopes that are present in several different HCV strains. Protease activity is provided by NS3 and NS3/NS4A proteins digesting the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide at 15 the appropriate cleavage sites to release polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B. Self- processing of the Met-NS3-NS4A-NS4B-NS5A-NS5B generates polypeptides that approximate naturally occurring HCV polypeptides.

Based on the guidance provided herein a sufficiently strong immune 20 response can be generated to achieve beneficial effects in a patient. The provided guidance includes information concerning HCV sequence selection, vector selection, vector production, combination treatment, and administration.

I. HCV SEQUENCES

25 A variety of different nucleic acid sequences can be used as a vaccine component to supply a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to a cell or as an intermediate to produce vaccine components. The starting point for obtaining suitable nucleic acid sequences are preferably naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences modified to produce an inactive NS5B.

30 The use of a HCV nucleic acid sequence providing HCV non-structural antigens to generate a CMI response is mentioned by Cho *et al.*, *Vaccine* 17:1136-1144, 1999, Paliard *et al.*, International Publication Number WO 01/30812 (not admitted to be prior art to the claimed invention), and Coit *et al.*, International Publication Number WO 01/38360 (not admitted to be prior art to the claimed 35 invention). Such references fail to describe, for example, a polypeptide that processes

itself to produce an inactive NS5B, and the particular combinations of HCV sequences and delivery vehicles employed herein.

Modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequence can be produced by altering the encoding nucleic acid.

5. Alterations can be performed to create deletions, insertions and substitutions.

Small modifications can be made in NS5B to produce an inactive polymerase by targeting motifs essentially for replication. Examples of motifs critical for NS5B activity and modifications that can be made to produce an inactive NS5B are described by Lohmann *et al.*, *Journal of Virology* 71:8416-8426, 1997, and

10 Kolykhalov *et al.*, *Journal of Virology* 74:2046-2051, 2000.

Additional factors to take into account when producing modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide include maintaining the ability to self-process and maintaining T cell antigens. The ability of the HCV polypeptide to process itself is determined to a large extent by a functional NS3 protease. Modifications that maintain NS3 activity protease activity can be obtained by taking into account the NS3 protein, NS4A which serves as a cofactor for NS3, and NS3 protease recognition sites present within the NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

Different modifications can be made to naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences to produce polypeptides able to elicit a broad range of T cell responses. Factors influencing the ability of a polypeptide to elicit a broad T cell response include the preservation or introduction of HCV specific T cell antigen regions and prevalence of different T cell antigen regions in different HCV isolates.

25 Numerous examples of naturally occurring HCV isolates are well known in the art. HCV isolates can be classified into the following six major genotypes comprising one or more subtypes: HCV-1/(1a,1b,1c), HCV-2/(2a,2b,2c), HCV-3/(3a,3b,10a), HCV-4/(4a), HCV-5/(5a) and HCV-6/(6a,6b,7b,8b,9a,11a). (Simmonds, *J. Gen. Virol.*, 693-712, 2001.) Examples of particular HCV sequences 30 such as HCV-BK, HCV-J, HCV-N, HCV-H, have been deposited in GenBank and described in various publications. (See, for example, Chamberlain *et al.*; *J. Gen. Virol.*, 1341-1347, 1997.)

35 HCV T cell antigens can be identified by, for example, empirical experimentation. One way of identifying T cell antigens involves generating a series of overlapping short peptides from a longer length polypeptide and then screening the

T-cell populations from infected patients for positive clones. Positive clones are activated/primed by a particular peptide. Techniques such as IFN γ -ELISPOT, IFN γ -Intracellular staining and bulk CTL assays can be used to measure peptide activity. Peptides thus identified can be considered to represent T-cell epitopes of the 5 respective pathogen.

HCV T cell antigen regions from different HCV isolates can be introduced into a single sequence by, for example, producing a hybrid NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing regions from two or more naturally occurring sequences. Such a hybrid can contain additional modifications, which 10 preferably do not reduce the ability of the polypeptide to produce an HCV CMI response.

15 The ability of a modified Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to process itself and produce a CMI response can be determined using techniques described herein or well known in the art. Such techniques include the use of IFN γ -ELISPOT, IFN γ -Intracellular staining and bulk CTL assays to measure a HCV specific CMI response.

A. Met-NS3-NS4A-NS4B-NS5A-NS5B Sequences

20 SEQ. ID. NO. 1 provides a preferred Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. SEQ. ID. NO. 1 contains a large number of HCV specific T cell antigens that are present in several different HCV isolates. SEQ. ID. NO. 1 is similar to the NS3-NS4A-NS4B-NS5A-NS5B portion of the HCV BK strain nucleotide sequence (GenBank accession number M58335).

25 In SEQ. ID. NO. 1 anchor positions important for recognition by MHC class I molecules are conserved or represent conservative substitutions for 18 out of 20 known T-cell epitopes in the NS3-NS4A-NS4B-NS5A-NS5B portion of HCV polyproteins. With respect to the remaining two known T-cell epitopes, one has a non-conservative anchor substitution in SEQ. ID. NO. 1 that may still be recognized by a different HLA supertype and one epitope has one anchor residue not conserved. 30 HCV T-cell epitopes are described in Chisari *et al.*, *Curr. Top. Microbiol. Immunol.*, 242:299-325, 2000, and Lechner *et al.* *J. Exp. Med.* 9:1499-1512, 2000.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequence and SEQ. ID. NO. 1 include the introduction of a methionine at the 5' end and the presence of modified NS5B active site residues in SEQ. ID. NO. 1.

The modification replaces GlyAspAsp with AlaAlaGly (residues 1711-1713) to inactivate NS5B.

The encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide preferably has an amino acid sequence substantially similar to SEQ. ID. NO. 1. In 5 different embodiments, the encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has an amino acid identity to SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or differs from SEQ. ID. NO. 1 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

10 Amino acid differences between a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide and SEQ. ID. NO. 1 are calculated by determining the minimum number of amino acid modifications in which the two sequences differ. Amino acid modifications can be deletions, additions, substitutions or any combination thereof.

15 Amino acid sequence identity is determined by methods well known in the art that compare the amino acid sequence of one polypeptide to the amino acid sequence of a second polypeptide and generate a sequence alignment. Amino acid identity is calculated from the alignment by counting the number of aligned residue pairs that have identical amino acids.

Methods for determining sequence identity include those described by 20 Schuler, G.D. in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouellette, B.F.F., eds., John Wiley & Sons, Inc, 2001; Yona, *et al.*, in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001).

25 Methods to determine amino acid sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul *et al.*, *J. Mol. Biol.* 215(3):403-10, 1990), and FASTA (Pearson, *Methods in Enzymology* 183:63-98, 1990, R.F. Doolittle, ed.).

30 In an embodiment of the present invention sequence identity between two polypeptides is determined using the GAP program (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman, *et al.*, *J. Mol. Biol.* 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two 35 sequences and creates a global alignment that maximizes the number of matched

residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment.

Default program parameters for polypeptide comparisons using GAP are the

5 BLOSUM62 (Henikoff *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:10915-10919, 1992) amino acid scoring matrix (MATrix=blosum62.cmp), a gap creation parameter (GAPweight=8) and a gap extension parameter (LENgthweight=2).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptides in addition to being substantially similar to SEQ. ID. NO. 1 across their 10 entire length produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 1. The corresponding regions in SEQ. ID. NO. 1 are provided as follows: Met-NS3 amino acids 1-632; NS4A amino acids 633-686; NS4B amino acids 687-947; NS5A amino acids 948-1394; and NS5B amino acids 1395-1985.

15 In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B region has an amino acid identity to the corresponding region in SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or an amino acid difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

20 Amino acid modifications to SEQ. ID. NO. 1 preferably maintain all or most of the T-cell antigen regions. Differences in naturally occurring amino acids are due to different amino acid side chains (R groups). An R group affects different properties of the amino acid such as physical size, charge, and hydrophobicity. Amino acids can be divided into different groups as follows: neutral and hydrophobic 25 (alanine, valine, leucine, isoleucine, proline, tryptophan, phenylalanine, and methionine); neutral and polar (glycine, serine, threonine, tryosine, cysteine, asparagine, and glutamine); basic (lysine, arginine, and histidine); and acidic (aspartic acid and glutamic acid).

30 Generally, in substituting different amino acids it is preferable to exchange amino acids having similar properties. Substituting different amino acids within a particular group, such as substituting valine for leucine, arginine for lysine, and asparagine for glutamine are good candidates for not causing a change in polypeptide tertiary structure.

35 Starting with a particular amino acid sequence and the known degeneracy of the genetic code, a large number of different encoding nucleic acid

sequences can be obtained. The degeneracy of the genetic code arises because almost all amino acids are encoded by different combinations of nucleotide triplets or "codons". The translation of a particular codon into a particular amino acid is well known in the art (see, e.g., Lewin *GENES IV*, p. 119, Oxford University Press, 1990).

5 Amino acids are encoded by codons as follows:

A=Ala=Alanine: codons GCA, GCC, GCG, GCU

C=Cys=Cysteine: codons UGC, UGU

D=Asp=Aspartic acid: codons GAC, GAU

E=Glu=Glutamic acid: codons GAA, GAG

10 F=Phe=Phenylalanine: codons UUC, UUU

G=Gly=Glycine: codons GGA, GGC, GGG, GGU

H=His=Histidine: codons CAC, CAU

I=Ile=Isoleucine: codons AUA, AUC, AUU

K=Lys=Lysine: codons AAA, AAG

15 L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU

M=Met=Methionine: codon AUG

N=Asn=Asparagine: codons AAC, AAU

P=Pro=Proline: codons CCA, CCC, CCG, CCU

Q=Gln=Glutamine: codons CAA, CAG

20 R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU

S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU

T=Thr=Threonine: codons ACA, ACC, ACG, ACU

V=Val=Valine: codons GUA, GUC, GUG, GUU

W=Trp=Tryptophan: codon UGG

25 Y=Tyr=Tyrosine: codons UAC, UAU.

Nucleic acid sequences can be optimized in an effort to enhance expression in a host. Factors to be considered include C:G content, preferred codons, and the avoidance of inhibitory secondary structure. These factors can be combined in different ways in an attempt to obtain nucleic acid sequences having enhanced expression in a particular host. (See, for example, Donnelly *et al.*, International Publication Number WO 97/47358.)

The ability of a particular sequence to have enhanced expression in a particular host involves some empirical experimentation. Such experimentation involves measuring expression of a prospective nucleic acid sequence and, if needed, altering the sequence.

B. Encoding Nucleotide Sequences

SEQ. ID. NOS. 2 and 3 provide two examples of nucleotide sequences encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. The coding sequence of SEQ. ID. NO. 2 is similar (99.4% nucleotide sequence identity) to the NS3-NS4A-NS4B-NS5A-NS5B region of the naturally occurring HCV-BK sequence (GenBank accession number M58335). SEQ. ID. NO. 3 is a codon-optimized version of SEQ. ID. NO. 2. SEQ. ID. NOS. 2 and 3 have a nucleotide sequence identity of 78.3%.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide (GenBank accession number M58335) and SEQ. ID. NO. 2, include SEQ. ID. NO. 2 having a ribosome binding site, an ATG methionine codon, a region coding for a modified NS5B catalytic domain, a TAAA stop signal and an additional 30 nucleotide differences. The modified catalytic domain codes for a AlaAlaGly (residues 1711-1713) instead of GlyAspAsp to inactivate NS5B.

A nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is preferably substantially similar to the SEQ. ID. NO. 2 coding region. In different embodiments, the nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has a nucleotide sequence identify to the SEQ. ID. NO. 2 coding region of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or differs from SEQ. ID. NO. 2 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

Nucleotide differences between a sequence coding Met-NS3-NS4A-NS4B-NS5A-NS5B and the SEQ. ID. NO. 2 coding region are calculated by determining the minimum number of nucleotide modifications in which the two sequences differ. Nucleotide modifications can be deletions, additions, substitutions or any combination thereof.

Nucleotide sequence identity is determined by methods well known in the art that compare the nucleotide sequence of one sequence to the nucleotide sequence of a second sequence and generate a sequence alignment. Sequence identity is determined from the alignment by counting the number of aligned positions having identical nucleotides.

Methods for determining nucleotide sequence identity between two polynucleotides include those described by Schuler, in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouelette, B.F.F.,

eds., John Wiley & Sons, Inc, 2001; Yona *et al.*, in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001). Methods to determine nucleotide sequence 5 identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul *et al.*, *J. Mol. Biol.* 215(3):403-10; 1990), and FASTA (Pearson, W.R., *Methods in Enzymology* 183:63-98, 1990, R.F. Doolittle, ed.).

In an embodiment of the present invention, sequence identity between 10 two polynucleotides is determined by application of GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman *et al.*, *J. Mol. Biol.* 48:443-453, 1970.) GAP considers all possible alignments and gap positions between 15 two sequences and creates a global alignment that maximizes the number of matched residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment. Default program parameters for polynucleotide comparisons using GAP are the 20 nwsgapdna.cmp scoring matrix (MATrix=nwsgapdna.cmp), a gap creation parameter (GAPweight=50) and a gap extension parameter (LENgthweight=3).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequences in addition to being substantially similar across its entire length, produce 25 individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 2. The corresponding coding regions in SEQ. ID. NO. 2 are provided as follows: Met-NS3, nucleotides 7-1902; NS4A nucleotides 1903-2064; NS4B nucleotides 2065-2847; NS5A nucleotides 2848-4188; NS5B nucleotides 4189-5661.

In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B 30 encoding region has a nucleotide sequence identity to the corresponding region in SEQ. ID. NO. 2 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference to SEQ. ID. NO. 2 of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

C. Gene Expression Cassettes

A gene expression cassette contains elements needed for polypeptide expression. Reference to "polypeptide" does not provide a size limitation and includes protein. Regulatory elements present in a gene expression cassette generally 5 include: (a) a promoter transcriptionally coupled to a nucleotide sequence encoding the polypeptide, (b) a 5' ribosome binding site functionally coupled to the nucleotide sequence, (c) a terminator joined to the 3' end of the nucleotide sequence, and (d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence. Additional regulatory elements useful for enhancing or regulating gene expression or polypeptide 10 processing may also be present.

Promoters are genetic elements that are recognized by an RNA polymerase and mediate transcription of downstream regions. Preferred promoters are strong promoters that provide for increased levels of transcription. Examples of 15 strong promoters are the immediate early human cytomegalovirus promoter (CMV), and CMV with intron A. (Chapman *et al.*, *Nucl. Acids Res.* 19:3979-3986, 1991.) Additional examples of promoters include naturally occurring promoters such as the EF1 alpha promoter, the murine CMV promoter, Rous sarcoma virus promoter, and 20 SV40 early/late promoters and the β -actin promoter; and artificial promoters such as a synthetic muscle specific promoter and a chimeric muscle-specific/CMV promoter (Li *et al.*, *Nat. Biotechnol.* 17:241-245, 1999, Hagstrom *et al.*, *Blood* 95:2536-2542, 2000).

The ribosome binding site is located at or near the initiation codon. Examples of preferred ribosome binding sites include CCACCAU^GG, 25 CCGCCAU^GG, and ACCAU^GG, where AUG is the initiation codon. (Kozak, *Cell* 44:283-292, 1986). Another example of a ribosome binding site is GCCACCAU^GG (SEQ. ID. NO. 12).

The polyadenylation signal is responsible for cleaving the transcribed RNA and the addition of a poly (A) tail to the RNA. The polyadenylation signal in 30 higher eukaryotes contains an AAUAAA sequence about 11-30 nucleotides from the polyadenylation addition site. The AAUAAA sequence is involved in signaling RNA cleavage. (Lewin, *Genes IV*, Oxford University Press, NY, 1990.) The poly (A) tail is important for the mRNA processing.

Polyadenylation signals that can be used as part of a gene expression 35 cassette include the minimal rabbit β -globin polyadenylation signal and the bovine growth hormone polyadenylation (BGH). (Xu *et al.*, *Gene* 272:149-156, 2001, Post *et*

al., U.S. Patent U. S. 5,122,458.) Additional examples include the Synthetic Polyadenylation Signal (SPA) and SV40 polyadenylation signal. The SPA sequence is as follows: AAUAAAAGAUCUUUUUUCAUUAGAUCUGUGUG, UUGGUUUUUUGUGUG (SEQ. ID. NO. 13).

5 Examples of additional regulatory elements useful for enhancing or regulating gene expression or polypeptide processing that may be present include an enhancer, a leader sequence and an operator. An enhancer region increases transcription. Examples of enhancer regions include the CMV enhancer and the SV40 enhancer. (Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, Xu, *et al.*, *Gene* 272:149-156, 2001.) An enhancer region can be associated with a promoter.

10 A leader sequence is an amino acid region on a polypeptide that directs the polypeptide into the proteasome. Nucleic acid encoding the leader sequence is 5' of a structural gene and is transcribed along the structural gene. An example of a leader sequences is tPA.

15 An operator sequence can be used to regulate gene expression. For example, the Tet operator sequence can be used to repress gene expression.

II. THERAPEUTIC VECTORS

20 Nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide can be introduced into a patient using vectors suitable for therapeutic administration. Suitable vectors can deliver nucleic acid into a target cell without causing an unacceptable side effect.

25 Cellular expression is achieved using a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide. The gene expression cassette contains regulatory elements for producing and processing a sufficient amount of nucleic acid inside a target cell to achieve a beneficial effect.

30 Examples of vectors that can be used for therapeutic applications include first and second generation adenovectors, helper dependent adenovectors, adeno-associated viral vectors, retroviral vectors, alpha virus vectors, Venezuelan Equine Encephalitis virus vector, and plasmid vectors. (Hitt, *et al.*, *Advances in Pharmacology* 40:137-206, 1997, Johnston *et al.*, U.S. Patent No. 6,156,588; and Johnston *et al.*, International Publication Number WO 95/32733.) Preferred vectors for introducing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide into a subject are first generation adenoviral vectors and plasmid DNA vectors.

A. First Generation Adenovectors

First generation adenovector for expressing a gene expression cassette contain the expression cassette in an E1 and optionally E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements needed for adenoviral replication.

First generation adenovectors for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide contain a E1 and E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements needed for adenoviral replication. The combinations of deletions of the E1 and E3 regions are sufficiently large to accommodate a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

The adenovirus has a double-stranded linear genome with inverted terminal repeats at both ends. During viral replication, the genome is packaged inside a viral capsid to form a virion. The virus enters its target cell through viral attachment followed by internalization. (Hitt *et al.*, *Advances in Pharmacology* 40:137-206, 1997.)

Adenovectors can be based on different adenovirus serotypes such as those found in humans or animals. Examples of animal adenoviruses include bovine, porcine, chimp, murine, canine, and avian (CELO). Preferred adenovectors are based on human serotypes, more preferably Group B, C, or D serotypes. Examples of human adenovirus Group B, C, D, or E serotypes include types 2 ("Ad2"), 4 ("Ad4"), 5 ("Ad5"), 6 ("Ad6"), 24 ("Ad24"), 26 ("Ad26"), 34 ("Ad34") and 35 ("Ad35"). Adenovectors can contain regions from a single adenovirus or from two or more adenovirus.

In different embodiments adenovectors are based on Ad5, Ad6, or a combination thereof. Ad5 is described by Chroboczek, *et al.*, *J. Virology* 186:280-285, 1992. Ad6 is described in Figures 7A-7N. An Ad6 based vector containing Ad5 regions is described in the Example section provided below.

Adenovectors do not need to have their E1 and E3 regions completely removed. Rather, a sufficient amount the E1 region is removed to render the vector replication incompetent in the absence of the E1 proteins being supplied in *trans*; and the E1 deletion or the combination of the E1 and E3 deletions are sufficiently large enough to accommodate a gene expression cassette.

E1 deletions can be obtained starting at about base pair 342 going up to about base pair 3523 of Ad5, or a corresponding region from other adenoviruses.

Preferably, the deleted region involves removing a region from about base pair 450 to about base pair 3511 of Ad5, or a corresponding region from other adenoviruses. Larger E1 region deletions starting at about base pair 341 removes elements that facilitate virus packaging.

5 E3 deletions can be obtained starting at about base pair 27865 to about base pair 30995 of Ad5, or the corresponding region of other adenovectors.

Preferably the deletion region involves removing a region from about base pair 28134 up to about base pair 30817 of Ad5, or the corresponding region of other adenovectors.

10 The combination of deletions to the E1 region and optionally the E3 region should be sufficiently large so that the overall size of the recombinant genome containing the gene expression cassette does not exceed about 105% of the wild type adenovirus genome. For example, as recombinant adenovirus Ad5 genomes increase size above about 105% the genome becomes unstable. (Bett *et al.*, *Journal of Virology* 67:5911-5921, 1993.)

15 Preferably, the size of the recombinant adenovirus genome containing the gene expression cassette is about 85% to about 105% the size of the wild type adenovirus genome. In different embodiments, the size of the recombinant adenovirus genome containing the expression cassette is about 100% to about 105.2%, or about 100%, the size of the wild type genome.

20 Approximately 7,500 kb can be inserted into an adenovirus genome with a E1 and E3 deletion. Without any deletion, the Ad5 genome is 35,935 base pairs and the Ad6 genome is 35,759 base pairs.

25 Replication of first generation adenovectors can be performed by supplying the E1 gene products in *trans*. The E1 gene product can be supplied in *trans*, for example, by using cell lines that have been transformed with the adenovirus E1 region. Examples of cells and cells lines transformed with the adenovirus E1 region are HEK 293 cells, 911 cells, PERC.6TM cells, and transfected primary human amniocytes cells. (Graham *et al.*, *Journal of Virology* 36:59-72, 1977, Schiedner *et al.*, *Human Gene Therapy* 11:2105-2116, 2000, Fallaux *et al.*, *Human Gene Therapy* 9:1909-1917, 1998, Bout *et al.*, U.S. Patent No. 6,033,908.)

30 A Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette should be inserted into a recombinant adenovirus genome in the region corresponding to the deleted E1 region or the deleted E3 region. The expression cassette can have a parallel or anti-parallel orientation. In a parallel orientation the transcription direction

of the inserted gene is the same direction as the deleted E1 or E3 gene. In an anti-parallel orientation transcription the opposite strand serves as a template and the transcription direction is in the opposite direction.

5 In an embodiment of the present invention the adenovector has a gene expression cassette inserted in the E1 deleted region. The vector contains:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6 joined to the fourth region.

In another embodiment of the present invention the adenovector has an expression cassette inserted in the E3 deleted region. The vector contains:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;

- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about 5 base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

In preferred different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first 10 region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

B. DNA Plasmid Vectors

15 DNA vaccine plasmid vectors contain a gene expression cassette along with elements facilitating replication and preferably vector selection. Preferred elements provide for replication in non-mammalian cells and a selectable marker. The vectors should not contain elements providing for replication in human cells or for integration into human nucleic acid.

20 The selectable marker facilitates selection of nucleic acids containing the marker. Preferred selectable markers are those that confer antibiotic resistance. Examples of antibiotic selection genes include nucleic acid encoding resistance to ampicillin, neomycin, and kanamycin.

25 Suitable DNA vaccine vectors can be produced starting with a plasmid containing a bacterial origin of replication and a selectable marker. Examples of bacterial origins of replication providing for higher yields include the ColE1 plasmid-derived bacterial origin of replication. (Donnelly *et al.*, *Annu. Rev. Immunol.* 15:617-648, 1997.)

30 The presence of the bacterial origin of replication and selectable marker allows for the production of the DNA vector in a bacterial strain such as *E. coli*. The selectable marker is used to eliminate bacteria not containing the DNA vector.

III. AD6 RECOMBINANT NUCLEIC ACID

Ad6 recombinant nucleic acid comprises an Ad6 region substantially similar to an Ad6 region found in SEQ. ID. NO. 8, and a region not present in Ad6 nucleic acid. Recombinant nucleic acid comprising Ad6 regions have different uses such as in producing different Ad6 regions, as intermediates in the production of Ad6 based vectors, and as a vector for delivering a recombinant gene.

As depicted in Figure 9, the genomic organization of Ad6 is very similar to the genomic organization of Ad5. The homology between Ad5 and Ad6 is approximately 98%.

In different embodiments, the Ad6 recombinant nucleic acid comprises a nucleotide region substantially similar to E1A, E1B, E2B, E2A, E3, E4, L1, L2, L3, or L4, or any combination thereof. A substantially similar nucleic acid region to an Ad6 region has a nucleotide sequence identity of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides. Techniques and embodiments for determining substantially similar nucleic acid sequences are described in Section I.B. *supra*.

Preferably, the recombinant Ad6 nucleic acid contains an expression cassette coding for a polypeptide not found in Ad6. Examples of expression cassettes include those coding for HCV regions and those coding for other types of polypeptides.

Different types of adenoviral vectors can be produced incorporating different amounts of Ad6, such as first and second generation adenovectors. As noted in Section II.A. *supra*, first generation adenovectors are defective in E1 and can replicate when E1 is supplied *in trans*.

Second generation adenovectors contain less adenoviral genome than first generation vectors and can be used in conjugation with complementing cell lines and/or helper vectors supplying adenoviral proteins. Second generation adenovectors are described in different references such as Russell, *Journal of General Virology* 81:2573-2604, 2000; Hitt *et al.*, 1997, Human Ad vectors for Gene Transfer, *Advances in Pharmacology*, Vol 40 Academic Press.

In an embodiment of the present invention, the Ad6 recombinant nucleic acid is an adenovirus vector defective in E1 that is able to replicate when E1 is

supplied in *trans*. Expression cassettes can be inserted into a deleted E1 region and/or a deleted E3 region.

An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E1 region comprises or consists of:

- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
- c) a second adenovirus region from about base pair 3511 to about 10 base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- 15 e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about base pair 30788 corresponding to Ad6, joined to the third region;
- f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base 20 pair 33784 corresponding to Ad6, wherein the fifth region is joined to the fourth region if the fourth region is present, or the fifth is joined to the third region if the fourth region is not present; and
- g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base 25 pair 35759 corresponding to Ad6, joined to the fifth region;
 wherein at least one Ad6 region is present.

In different embodiments of the invention, all of the regions are from Ad6; all of the regions except for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth, and fifth regions are from Ad6.

- 30 An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E3 region comprises or consists of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;

5 c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;

d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;

10 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and

f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region;

15 wherein at least one Ad6 region is present.

In different embodiment of the invention, all of the regions are from Ad6; all of the regions except for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth and fifth regions are from Ad6.

IV. VECTOR PRODUCTION

25 Vectors can be produced using recombinant nucleic acid techniques such as those involving the use of restriction enzymes, nucleic acid ligation, and homologous recombination. Recombinant nucleic acid techniques are well known in the art. (Ausubel, *Current Protocols in Molecular Biology*, John Wiley, 1987-1998, and Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, 1989.)

Intermediate vectors are used to derive a therapeutic vector or to transfer an expression cassette or portion thereof from one vector to another vector. Examples of intermediate vectors include adenovirus genome plasmids and shuttle vectors.

Useful elements in an intermediate vector include an origin of replication, a selectable marker, homologous recombination regions, and convenient restriction sites. Convenient restriction sites can be used to facilitate cloning or release of a nucleic acid sequence.

Homologous recombination regions provide nucleic acid sequence regions that are homologous to a target region in another nucleic acid molecule. The homologous regions flank the nucleic acid sequence that is being inserted into the target region. In different embodiments homologous regions are preferably about 150 to 600 nucleotides in length, or about 100 to 500 nucleotides in length.

An embodiment of the present invention describes a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette, a selectable marker, a bacterial origin of replication, a first adenovirus homology region and a second adenovirus homologous region that target the expression cassette to insert in or replace an E1 region. The first and second homology regions flank the expression cassette. The first homology region contains at least about 100 base pairs substantially homologous to at least the right end (3' end) of a wild-type adenovirus region from about base pairs 4-450. The second homology contains at least about 100 base pairs substantially homologous to at least the left end (5' end) of Ad5 from about base pairs 3511-5792, or the corresponding region from another adenovirus.

Reference to "substantially homologous" indicates a sufficient degree of homology to specifically recombine with a target region. In different embodiments substantially homologous refers to at least 85%, at least 95%, or 100% sequence identity. Sequence identity can be calculated as described in Section I.B. *supra*.

One method of producing adenovectors is through the creation of an adenovirus genome plasmid containing an expression cassette. The pre-Adenovirus plasmid contains all the adenovirus sequences needed for replication in the desired complementing cell line. The pre-Adenovirus plasmid is then digested with a restriction enzyme to release the viral ITR's and transfected into the complementing cell line for virus rescue. The ITR's must be released from plasmid sequences to allow replication to occur. Adenovector rescue results in the production of an adenovector containing the expression cassette.

A. Adenovirus Genome Plasmids

Adenovirus genome plasmids contain an adenovector sequence inside a longer-length plasmid (which may be a cosmid). The longer-length plasmid may contain additional elements such as those facilitating growth and selection in eukaryotic or bacterial cells depending upon the procedures employed to produce and maintain the plasmid. Techniques for producing adenovirus genome plasmids include those involving the use of shuttle vectors and homologous recombination, and those

involving the insertion of a gene expression cassette into an adenovirus cosmid. (Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, Danthinne *et al.*, *Gene Therapy* 7:1707-1714, 2000.)

Adenovirus genome plasmids preferably have a gene expression cassette inserted into a E1 or E3 deleted region. In an embodiment of the present invention, the adenovirus genome plasmid contains a gene expression cassette inserted in the E1 deleted region, an origin of replication, a selectable marker, and the recombinant adenovirus region is made up of:

- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- 10 b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 15 5541 corresponding to Ad6, joined to the expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 20 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region;
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 25 35759 corresponding to Ad6, joined to the fourth region, and
- g) an optionally present E3 region corresponding to all or part of the E3 region present in Ad5 or Ad6, which may be present for smaller inserts taking into account the overall size of the desired adenovector.

In another embodiment of the present invention the recombinant adenovirus genome plasmid has the gene expression cassette inserted in the E3 deleted region. The vector contains an origin of replication, a selectable marker, and the following:

- 30 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
- c) a third adenovirus region from about base pair 5549 to about 5 base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- d) the gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
- e) a fourth adenovirus region from about base pair 30818 to about 10 base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

15 In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

20

An embodiment of the present invention describes a method of making an adenovector involving a homologous recombination step to produce a adenovirus genome plasmid and an adenovirus rescue step. The homologous recombination step involves the use of a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette flanked by adenovirus homology regions. The adenovirus homology regions target the expression cassette into either the E1 or E3 deleted region.

25 In an embodiment of the present invention concerning the production of an adenovirus genome plasmid, the gene expression cassette is inserted into a vector comprising: a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6; a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the second region; a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding 30

35 to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6,

joined to the second region; a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and a fifth adenovirus region from about 33967 to about 35935 corresponding to Ad5 or from 5 about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region. The adenovirus genome plasmid should contain an origin of replication and a selectable marker, and may contain all or part of the Ad5 or Ad6 E3 region.

In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the 10 first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

15 **B. Adenovector Rescue**

An adenovector can be rescued from a recombinant adenovirus genome plasmid using techniques known in the art or described herein. Examples of techniques for adenovirus rescue well known in the art are provided by Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, and Danthinne *et al.*, *Gene Therapy* 20 7:1707-1714, 2000.

A preferred method of rescuing an adenovector described herein involves boosting adenoviral replication. Boosting adenoviral replication can be performed, for example, by supplying adenoviral functions such as E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 on a 25 separate plasmid. Example 10 *infra*. illustrates the boosting of adenoviral replication to rescue an adenovector containing a codon optimized Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette.

V. PARTIAL-OPTIMIZED HCV ENCODING SEQUENCES

30 Partial optimization of HCV polyprotein encoding nucleic acid provides for a lesser amount of codons optimized for expression in a human than complete optimization. The overall objective is to provide the benefits of increased expression due to codon optimization, while facilitating the production of an adenovector containing HCV polyprotein encoding nucleic acid having optimized 35 codons.

Complete optimization of an HCV polyprotein encoding sequence provides the most frequently observed human codon for each amino acid. Complete optimization can be performed using codon frequency tables well known in the art and using programs such as the BACKTRANSLATE program (Wisconsin Package 5 version 10, Genetics Computer Group, GCG, Madison, Wisc.).

Partial optimization can be performed on an entire HCV polyprotein encoding sequence that is present (e.g., NS3-NS5B), or one or more local regions that are present. In different embodiments the GC content for the entire HCV encoded polyprotein that is present is no greater than at least about 65%; and the GC content for one or more 10 local regions is no greater than about 70%.

Local regions are regions present in HCV encoding nucleic acid, and can vary in size. For example, local regions can be about 60, about 70, about 80, about 90 or about 100 nucleotides in length.

Partial optimization can be achieved by initially constructing an HCV 15 encoding polyprotein sequence to be partially optimized based on a naturally occurring sequence. Alternatively, an optimized HCV encoding sequence can be used as basis of comparison to produce a partial optimized sequence.

VI. HCV COMBINATION TREATMENT

20 The HCV Met-NS3-NS4A-NS4B-NS5A-NS5B vaccine can be used by itself to treat a patient, can be used in conjunction with other HCV therapeutics, and can be used with agents targeting other types of diseases. Additional therapeutics include additional therapeutic agents to treat HCV and diseases having a high prevalence in HCV infected persons. Agents targeting other types of disease include 25 vaccines directed against HIV and HBV.

Additional therapeutics for treating HCV include vaccines and non-vaccine agents. (Zein, *Expert Opin. Investig. Drugs* 10:1457-1469, 2001.) Examples of additional HCV vaccines include vaccines designed to elicit an immune response against an HCV core antigen and the HCV E1, E2 or p7 region. Vaccine components 30 can be naturally occurring HCV polypeptides, HCV mimotope polypeptides or nucleic acid encoding such polypeptides.

HCV mimotope polypeptides contain HCV epitopes, but have a different sequence than a naturally occurring HCV antigen. A HCV mimotope can be fused to a naturally occurring HCV antigen. References describing techniques for 35 producing mimotopes in general and describing different HCV mimotopes are

provided in Felici *et al.* U.S. Patent No. 5,994,083 and Nicosia *et al.*, International Application Number WO 99/60132.

VII. PHARMACEUTICAL ADMINISTRATION

5 HCV vaccines can be formulated and administered to a patient using the guidance provided herein along with techniques well known in the art. Guidelines for pharmaceutical administration in general are provided in, for example, *Modern Vaccinology*, Ed. Kurstak, Plenum Med. Co. 1994; *Remington's Pharmaceutical Sciences 18th Edition*, Ed. Gennaro, Mack Publishing, 1990; and *Modern Pharmaceutics 2nd Edition*, Eds. Bunker and Rhodes, Marcel Dekker, Inc., 1990, each of which are hereby incorporated by reference herein.

10 HCV vaccines can be administered by different routes such as intravenous, intraperitoneal, subcutaneous, intramuscular, intradermal, impression through the skin, or nasal. A preferred route is intramuscular.

15 Intramuscular administration can be performed using different techniques such as by injection with or without one or more electric pulses. Electric mediated transfer can assist genetic immunization by stimulating both humoral and cellular immune responses.

20 Vaccine injection can be performed using different techniques, such as by employing a needle or a needless injection system. An example of a needless injection system is a jet injection device. (Donnelly *et al.*, International Publication Number WO 99/52463.)

A. Electrically Mediated Transfer

25 Electrically mediated transfer or Gene Electro-Transfer (GET) can be performed by delivering suitable electric pulses after nucleic acid injection. (See Mathiesen, International Publication Number WO 98/43702). Plasmid injection and electroporation can be performed using stainless needles. Needles can be used in couples, triplets or more complex patterns. In one configuration the needles are 30 soldered on a printed circuit board that is a mechanical support and connects the needles to the electrical field generator by means of suitable cables.

35 The electrical stimulus is given in the form of electrical pulses. Pulses can be of different forms (square, sinusoidal, triangular, exponential decay) and different polarity (monopolar or positive or negative polarity, bipolar). Pulses can be delivered either at constant voltage or constant current modality.

Different patterns of electric treatment can be used to introduce nucleic acid vaccines including HCV and other nucleic acid vaccines into a patient. Possible patterns of electric treatment include the following:

5 Treatment 1: 10 trains of 1000 square bipolar pulses delivered every other second, pulse length 0.2 msec/phase, frequency 1000 Hz, constant voltage mode, 45 Volts/phase, floating current.

10 Treatment 2: 2 trains of 100 square bipolar pulses delivered every other second, pulse length 2 msec/phase, frequency 100 Hz, constant current mode, 100 mA/phase, floating voltage.

15 Treatment 3: 2 trains of bipolar pulses at a pulse length of about 2 msec/phase, for a total length of about 3 seconds, where the actual current going through the tissue is fixed at about 50 mA.

Electric pulses are delivered through an electric field generator. A suitable generator can be composed of three independent hardware elements assembled in a common chassis and driven by a portable PC which runs the driving program. The software manages both basic and accessory functions. The elements of the device are: (1) signal generator driven by a microprocessor, (2) power amplifier and (3) digital oscilloscope.

20 The signal generator delivers signals having arbitrary frequency and shape in a given range under software control. The same software has an interactive editor for the waveform to be delivered. The generator features a digitally controlled current limiting device (a safety feature to control the maximal current output). The power amplifier can amplify the signal generated up to +/- 150 V. The oscilloscope is digital and is able to sample both the voltage and the current being delivered by the 25 amplifier.

B. Pharmaceutical Carriers

30 Pharmaceutically acceptable carriers facilitate storage and administration of a vaccine to a subject. Examples of pharmaceutically acceptable carriers are described herein. Additional pharmaceutical acceptable carriers are well known in the art.

35 Pharmaceutically acceptable carriers may contain different components such a buffer, normal saline or phosphate buffered saline, sucrose, salts and polysorbate. An example of a pharmaceutically acceptable carrier is follows: 2.5-10 mM TRIS buffer, preferably about 5 mM TRIS buffer; 25-100 mM NaCl; preferably

about 75 mM NaCl; 2.5-10% sucrose, preferably about 5% sucrose; 0.01 -2 mM MgCl₂; and 0.001%-0.01% polysorbate 80 (plant derived). The pH is preferably from about 7.0-9.0, more preferably about 8.0. A specific example of a carrier contains 5 mM TRIS, 75 mM NaCl, 5% sucrose, 1 mM MgCl₂, 0.005% polysorbate 80 at pH 5 8.0.

C. Dosing Regimes

10 Suitable dosing regimens can be determined taking into account the efficacy of a particular vaccine and factors such as age, weight, sex and medical condition of a patient; the route of administration; the desired effect; and the number of doses. The efficacy of a particular vaccine depends on different factors such as the ability of a particular vaccine to produce polypeptide that is expressed and processed in a cell and presented in the context of MHC class I and II complexes.

15 HCV encoding nucleic acid administered to a patient can be part of different types of vectors including viral vectors such as adenovector, and DNA plasmid vaccines. In different embodiments concerning administration of a DNA plasmid, about 0.1 to 10 mg of plasmid is administered to a patient, and about 1 to 5 mg of plasmid is administered to a patient. In different embodiments concerning administration of a viral vector, preferably an adenoviral vector, about 10⁵ to 10¹¹ 20 viral particles are administered to a patient, and about 10⁷ to 10¹⁰ viral particles are administered to a patient.

25 Viral vector vaccines and DNA plasmid vaccines may be administered alone, or may be part of a prime and boost administration regimen. A mixed modality priming and booster inoculation involves either priming with a DNA vaccine and boosting with viral vector vaccine, or priming with a viral vector vaccine and boosting with a DNA vaccine.

30 Multiple priming, for example, about to 2-4 or more may be used. The length of time between priming and boost may typically vary from about four months to a year, but other time frames may be used. The use of a priming regimen with a DNA vaccine may be preferred in situations where a person has a pre-existing anti-adenovirus immune response.

35 In an embodiment of the present invention, 1x10⁷ to 1x10¹² particles and preferably about 1x10¹⁰ to 1x10¹¹ particles of adenovector is administered directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

In another embodiment of the present invention initial vaccination is performed with a DNA vaccine directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

Agents such as interleukin-12, GM-CSF, B7-1, B7-2, IP10, Mig-1 can be coadministered to boost the immune response. The agents can be coadministered as proteins or through use of nucleic acid vectors.

D. Heterologous Prime-Boost

Heterologous prime-boost is a mixed modality involving the use of one type of viral vector for priming and another type of viral vector for boosting. The heterologous prime-boost can involve related vectors such as vectors based on different adenovirus serotypes and more distantly related viruses such adenovirus and poxvirus. The use of poxvirus and adenovirus vectors to protect mice against malaria is illustrated by Gilbert *et al.*, *Vaccine* 20:1039-1045, 2002.

Different embodiments concerning priming and boosting involve the following types of vectors expressing desired antigens such as Met-NS3-NS4A-NS4B-NS5A-NS5B: Ad5 vector followed by Ad6 vector; Ad6 vector followed by Ad5 vector; Ad5 vector followed by poxvirus vector; poxvirus vector followed by Ad5 vector; Ad6 vector followed by poxvirus vector; and poxvirus vector followed by Ad6 vector.

The length of time between priming and boosting typically varies from about four months to a year, but other time frames may be used. The minimum time frame should be sufficient to allow for an immunological rest. In an embodiment, this rest is for a period of at least 6 months. Priming may involve multiple priming with one type of vector, such as 2-4 primings.

Expression cassettes present in a poxvirus vector should contain a promoter either native to, or derived from, the poxvirus of interest or another poxvirus member. Different strategies for constructing and employing different types of poxvirus based vectors including those based on vaccinia virus, modified vaccinia virus, avipoxvirus, raccoon poxvirus, modified vaccinia virus Ankara, canarypoxviruses (such as ALVAC), fowlpoxviruses, cowpoxviruses, and NYVAC are well known in the art. (Moss, *Current Topics in Microbiology and Immunology* 158:25-38, 1982; Earl *et al.*, In *Current Protocols in Molecular Biology*, Ausubel *et al.* eds., New York: Greene Publishing Associates & Wiley Interscience; 1991:16.16.1-16.16.7, Child *et al.*, *Virology* 174(2):625-9, 1990; Tartaglia *et al.*,

Virology 188:217-232, 1992; U.S. Patent Nos., 4,603,112, 4,722,848, 4,769,330, 5,110,587, 5,174,993, 5,185,146, 5,266,313, 5,505,941, 5,863,542, and 5,942,235.

E. Adjuvants

5 HCV vaccines can be formulated with an adjuvant. Adjuvants are particularly useful for DNA plasmid vaccines. Examples of adjuvants are alum, AlPO₄, alhydrogel, Lipid-A and derivatives or variants thereof, Freund's incomplete adjuvant, neutral liposomes, liposomes containing the vaccine and cytokines, non-ionic block copolymers, and chemokines.

10 Non-ionic block polymers containing polyoxyethylene (POE) and polyxylpropylene (POP), such as POE-POP-POE block copolymers may be used as an adjuvant. (Newman *et al.*, *Critical Reviews in Therapeutic Drug Carrier Systems* 15:89-142, 1998.) The immune response of a nucleic acid can be enhanced using a non-ionic block copolymer combined with an anionic surfactant.

15 A specific example of an adjuvant formulation is one containing CRL-1005 (CytRx Research Laboratories), DNA, and benzylalkonium chloride (BAK). The formulation can be prepared by adding pure polymer to a cold (< 5°C) solution of plasmid DNA in PBS using a positive displacement pipette. The solution is then vortexed to solubilize the polymer. After complete solubilization of the polymer a clear solution is obtained at temperatures below the cloud point of the polymer (~6-7°C). Approximately 4 mM BAK is then added to the DNA/CRL-1005 solution in PBS, by slow addition of a dilute solution of BAK dissolved in PBS. The initial DNA concentration is approximately 6 mg/mL before the addition of polymer and BAK, and the final DNA concentration is about 5 mg/mL. After BAK addition the formulation is vortexed extensively, while the temperature is allowed to increase from ~ 2°C to above the cloud point. The formulation is then placed on ice to decrease the temperature below the cloud point. Then, the formulation is vortexed while the temperature is allowed to increase from ~2°C to above the cloud point. Cooling and mixing while the temperature is allowed to increase from ~2°C to above the cloud point is repeated several times, until the particle size of the formulation is about 200-500 nm, as measured by dynamic light scattering. The formulation is then stored on ice until the solution is clear, then placed in storage at -70°C. Before use, the formulation is allowed to thaw at room temperature.

F. Vaccine Storage

Adenovector and DNA vaccines can be stored using different types of buffers. For example, buffer A105 described in Example 9 *infra*. can be used to for vector storage.

5 Storage of DNA can be enhanced by removal or chelation of trace metal ions. Reagents such as succinic or malic acid, and chelators can be used to enhance DNA vaccine stability. Examples of chelators include multiple phosphate ligands and EDTA. The inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, can also be useful to prevent damage of DNA plasmid from free 10 radical production. Furthermore, the buffer type, pH, salt concentration, light exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine.

VII. EXAMPLES

15 Examples are provided below to further illustrate different features of the present invention. The examples also illustrate useful methodology for practicing the invention. These examples do not limit the claimed invention.

Example 1: Met-NS3-NS4A-NS4B-NS5A-NS5B Expression Cassettes

20 Different gene expression cassettes encoding HCV NS3-NS4A-NS4B-NS5A-NS5B were constructed based on a 1b subtype HCV BK strain. The encoded sequences had either (1) an active NS5B sequence ("NS"), (2) an inactive NS5B sequence ("NSmut"), (3) a codon optimized sequence with an inactive NS5B sequence ("NSOPTmut"). The expression cassettes also contained a CMV 25 promoter/enhancer and the BGH polyadenylation signal.

The NS nucleotide sequence (SEQ. ID. NO. 5) differs from HCV BK strain GenBank accession number M58335 by 30 out of 5952 nucleotides. The NS amino acid sequence (SEQ. ID. NO. 6) differs from the corresponding 1b genotype HCV BK strain by 7 out of 1984 amino acids. To allow for initiation of translation an 30 ATG codon is present at the 5' end of the NS sequence. A TGA termination sequence is present at the 3' end of the NS sequence.

The NSmut nucleotide sequence (SEQ. ID. NO. 2, Figure 2), is similar to the NS sequence. The differences between NSmut and NS include NSmut having an altered NS5B catalytic site; an optimal ribosome binding site at the 5' end; and a 35 TAAA termination sequence at the 3' end. The alterations in NS5B comprise bases

5138 to 5146, which encode amino acids 1711 to 1713. The alterations result in a change of amino acids GlyAspAsp into AlaAlaGly and creates an inactive form of the NS5B RNA-dependent RNA-polymerase NSSB.

The NSOPTmut sequence (SEQ. ID. NO. 3, Figure 3) was designed 5 based on the amino acid sequence encoded by NSmut. The NSmut amino acid sequence was back translated into a nucleotide sequence with the GCG (Wisconsin Package version 10, Genetics Computer Group, GCG, Madison, Wisc.) BACKTRANSLATE program. To generate a NSOPTmut nucleotide sequence where each amino acid is coded for by the corresponding most frequently observed human 10 codon, the program was run choosing as parameter the generation of the most probable nucleotide sequence and specifying the codon frequency table of highly expressed human genes (human_high.cod) available within the GCG Package as translation scheme.

15 Example 2: Generation pV1Jns plasmid with NS, NSmut or NSOPTmut Sequences
pV1Jns plasmids containing either the NS sequence, NSmut sequence or NSOPTmut sequences were generated and characterised as follows:

20 *pV1Jns Plasmid with the NS Sequence*
The coding region Met-NS3-NS4A-NS4B-NS5A and the coding region Met-NS3-NS4A-NS4B-NS5A-NS5B from a HCV BK type strain (Tomei *et al.*, *J. Virol.* 67:4017-4026, 1993) were cloned into pcDNA3 plasmid (Invitrogen), generating pcD3-5a and pcD3-5b vectors, respectively. Pcd3-5A was digested with Hind III, blunt-ended with Klenow fill-in and subsequently digested with Xba I, to 25 generate a fragment corresponding to the coding region of Met-NS3-NS4A-NS4B-NS5A. The fragment was cloned into pV1Jns-poly, digested with Bgl II blunt-ended with Klenow fill-in and subsequently digested with Xba I, generating pV1JnsNS3-5A.

25 pV1Jns-poly is a derivative of pV1JnsA plasmid (Montgomery *et al.*, *DNA and Cell Biol.* 12:777-783, 1993), modified by insertion of a polylinker 30 containing recognition sites for XbaI, PmeI, PacI into the unique BglII and NotI restriction sites. The pV1Jns plasmid with the NS sequence (pV1JnsNS3-5B) was obtained by homologous recombination into the bacterial strain BJ5183, co-transforming pV1JNS3-5A linearized with XbaI and NotI digestion and a PCR fragment containing approximately 200 bp of NS5A, NS5B coding sequence and

approximately 60 bp of the BGH polyadenylation signal. The resulting plasmid represents pV1Jns-NS.

pV1Jns-NS can be summarized as follows:

Bases	1 to 1881 of pV1JnsA
5 an additional then the	AGCTT
then the	Met-NS3-NS5B sequence (SEQ. ID. NO. 5)
an additional NO. 14)	wt TGA stop
10 Bases	TCTAGAGCGTTAAACCCCTTAATTAAGG (SEQ. ID.
	1912 to 4909 of pV1JnsA

pV1Jns Plasmid with the NSmut Sequence

The V1JnsNS3-5A plasmid was modified at the 5' of the NS3 coding sequence by addition of a full Kozak sequence. The plasmid (V1JNS3-5Akозак) was obtained by homologous recombination into the bacterial strain BJ5183, co-transforming V1JNS3-5A linearized by *Af*III digestion and a PCR fragment containing the proximal part of Intron A, the restriction site *Bgl*II, a full Kozak translation initiation sequence and part of the NS3 coding sequence.

The resulting plasmid (V1JNS3-5Akозак) was linearized with *Xba* I digestion and co-transformed into the bacterial strain BJ5183 with a PCR fragment, containing approximately 200 bp of NS5A, the NS5B mutated sequence, the strong translation termination TAAA and approximately 60 bp of the BGH polyadenylation signal. The PCR fragment was obtained by assembling two 22bp-overlapping fragments where mutations were introduced by the oligonucleotides used for their amplification. The resulting plasmid represents pV1Jns-NSmut.

pV1Jns-NSmut can be summarized as follows:

Bases	1 to 1882 of pV1JnsA
then the	kozak Met-NS3-NS5B(mut) TAAA sequence (SEQ. ID. NO. 2)
an additional	TCTAGA
30 Bases	1925 to 4909 of pV1JnsA

pV1Jns Plasmid with the NSOPTmut Sequence

The human codon-optimized synthetic gene (NSOPTmut) with mutated NS5B to abrogate enzymatic activity, full Kozak translation initiation sequence and a strong translation termination was digested with *Bam*HI and *Sall*

restriction sites present at the 5' and 3' end of the gene. The gene was then cloned into the BglII and SalI restriction sites present in the polylinker of pV1JnsA plasmid, generating pV1Jns-NSOPTmut.

pV1Jns-NSOPTmut can be summarized as follows:

5 Bases 1 to 1881 of pV1JnsA
an additional C
then kozak Met-NS3-NS5B(optmut) TAAA sequence (SEQ. ID. NO. 3)
an additional TTTAAATGTTAAC (SEQ. ID. NO. 15)
Bases 1905 to 4909 of pV1JnsA

10

Plasmids Characterization

Expression of HCV NS proteins was tested by transfection of HEK 293 cells, grown in 10% FCS/DMEM supplemented by L-glutamine (final 4 mM). Twenty-four hours before transfection, cells were plated in 6-well 35 mm diameter, to reach 90-95% confluence on the day of transfection. Forty nanograms of plasmid DNA (previously assessed as a non-saturating DNA amount) were co-transfected with 100 ng of pRSV-Luc plasmid containing the luciferase reporter gene under the control of Rous sarcoma virus promoter, using the LIPOFECTAMINE 2000 reagent. Cells were kept in a CO₂ incubator for 48 hours at 37 °C.

20 Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were normalized for Luciferase activity, and run in serial dilution on 10% SDS-acrylamide gel. Proteins were transferred on nitrocellulose and assayed with antibodies directed against NS3, NS5A and NS5B to assess strength of expression and correct proteolytic cleavage. Mock-transfected cells were used as a negative control.
25 Results from representative experiments testing pV1JnsNS, pV1JnsNSmut and pV1JnsNSOPTmut are shown in Figure 12.

Example 3: Mice Immunization with Plasmid DNA Vectors

The DNA plasmids pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPTmut were injected in different mice strains to evaluate their potential to elicit anti-HCV immune responses. Two different strains (Balb/C and C57Black6, N=9-10) were injected intramuscularly with 25 or 50 µg of DNA followed by electrical pluses. Each animal received two doses at three weeks interval.

30 Humoral immune response elicited in C57Black6 mice against the NS3 protein was measured in post dose two sera by ELISA on bacterially expressed NS3

protease domain. Antibodies specific for the tested antigen were detected in animals immunized with all three vectors with geometric mean titers (GMT) ranging from 94000 to 133000 (Tables 1-3).

5

Table 1: pV1jns-NS

Mice n.	1	2	3	4	5	6	7	8	9	GMT
	Titer	105466	891980	78799	39496	543542	182139	32351	95028	67800

10

Table 2: pV1jns-NSmut

Mice n.	11	12	13	14	15	16	17	18	19	20	GMT
	Titer	202981	55670	130786	49748	17672	174958	44304	37337	78182	193695

15

Table 3: pV1jns-NSOPTmut

Mice n.	21	22	23	24	25	26	27	28	29	30	GMT
	Titer	310349	43645	63496	82174	630778	297259	66861	146735	173506	77732

20

A T cell response was measured in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 25 µg of plasmid DNA. Quantitative ELIsot assay was performed to determine the number of IFN γ secreting

25 T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response was analyzed by the same assay using a 20mer peptide encompassing a CD8+ epitope for C57Black6 mice (pep1480).

Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELIsot assay. T cell response in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 50 µg of plasmid DNA, was

analyzed by the same ELIspot assay measuring the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence.

Spleen cells were prepared from immunized mice and re-suspended in
5 R10 medium (RPMI 1640 supplemented with 10% FCS, 2 mM L-Glutamine, 50 U/ml-50 μ g/ml Penicillin/Streptomycin, 10 mM Hepes, 50 μ M 2-mercapto-ethanol).
Multiscreen 96-well Filtration Plates (Millipore, Cat. No. MAIPS4510, Millipore Corporation, 80 Ashby Road Bedford, MA) were coated with purified rat anti-mouse
10 IFN γ antibody (PharMingen, Cat. No. 18181D, PharmiMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA). After overnight incubation, plates were washed with PBS 1X/0.005% Tween and blocked with 250 μ l/well of R10 medium.

15 Splenocytes from immunized mice were prepared and incubated for twenty-four hours in the presence or absence of 10 μ M peptide at a density of 2.5 X 10⁵/well or 5 X 10⁵/well. After extensive washing (PBS 1X/0.005% Tween), biotinylated rat anti-mouse IFN γ antibody (PharMingen, Cat. No. 18112D, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) was added and incubated overnight at 4° C. For development, streptavidin-AKP (PharMingen, Cat. No. 13043E, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) and 1-StepTM NBT-BCIP development solution (Pierce, Cat. No. 34042, Pierce, P.O. Box 117, Rockford, IL 61105 USA) were added.

20 Pools of 20mer overlapping peptides encompassing the entire sequence of the HCV BK strain NS3 to NS5B were used to reveal HCV-specific IFN γ -secreting T cells. Similarly a single 20mer peptide encompassing a CD8+ epitope for C57Black6 mice was used to detect CD8 response. Representative data from groups of C57Black6 and Balb/C mice (N=9-10) immunized with two injections of 25 or 50 μ g of plasmid vectors pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPTmut are shown in Figures 13A and 13B.

25 **30 Example 4: Immunization of Rhesus Macaques**

Rhesus macaques (N=3) were immunized by intramuscular injection with 5mg of plasmid pV1Jns-NSOPTmut in 7.5mg/ml CRL1005, Benzalkonium chloride 0.6 mM. Each animal received two doses in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by IFN- γ ELISPOT. This assay measures HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

5 The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5A, NS5B) were prepared for use in these assays to measure immune responses in HCV DNA and
10 adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

15

IFN γ ELISPOT

The IFN γ -ELISPOT assay provides a quantitative determination of HCV-specific T lymphocyte responses. PBMC are serially diluted and placed in microplate wells coated with anti-rhesus IFN- γ antibody (MD-1 U-Cytech). They are
20 cultured with a HCV peptide pool for 20 hours, resulting in the restimulation of the precursor cells and secretion of IFN- γ . The cells are washed away, leaving the secreted IFN bound to the antibody-coated wells in concentrated areas where the cells were sitting. The captured IFN is detected with biotinylated anti-rhesus IFN antibody (detector Ab U-Cytech) followed by alkaline phosphatase-conjugated streptavidin
25 (Pharmingen 13043E). The addition of insoluble alkaline phosphatase substrate results in dark spots in the wells at the sites where the cells were located, leaving one spot for each T cell that secreted IFN- γ .

The number of spots per well is directly related to the precursor frequency of antigen-specific T cells. Gamma interferon was selected as the cytokine
30 visualized in this assay (using species specific anti-gamma interferon monoclonal antibodies) because it is the most common, and one of the most abundant cytokines synthesized and secreted by activated T lymphocytes. For this assay, the number of spot forming cells (SFC) per million PBMCs is determined for samples in the

presence and absence (media control) of peptide antigens. Data from Rhesus macaques on PBMC from post dose two material are shown in Table 4.

Table 4

Pep pools	PV1J-NSOPTmut		
	21G	99C161	99C166
F (NS3p)	8	10	170
G (NS3h)	7	592	229
H (NS4)	3	14	16
I (NS5a)	5	71	36
L (NS5b)	14	23	11
M (NS5b)	3	35	8
DMSO	2	4	5

5 INF γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 5 mg DNA/dose in OPTIVAX/BAK of plasmid pV1Jns-NSOPTmut. Data are expressed as SFC7 10⁶ PBMC.

Example 5: Construction of Ad6 Pre-Adenovirus Plasmids

10 Ad6 pre-adenovirus plasmids were obtained as follows:

10

Construction of pAd6 E1-E3+ Pre-adenovirus Plasmid

15 An Ad6 based pre-adenovirus plasmid which can be used to generate first generation Ad6 vectors was constructed either taking advantage of the extensive sequence identity (approx. 98%) between Ad5 and Ad6 or containing only Ad6 regions. Homologous recombination was used to clone wtAd6 sequences into a bacterial plasmid.

20 A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid containing Ad5 and Ad6 regions is illustrated in Figure 10. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad5 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 33798 to 35935) and left (bp 1 to 341 and bp 3525 to 5767) end of the Ad5 genome separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. The ITR cassette contains a deletion of E1 sequences from

Ad5 342 to 3524. The Ad5 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral DNA in which recombination can occur.

Potential clones were screened by restriction analysis and one clone was selected as pAd6E1-E3+. This clone was then sequenced in its entirety. pAd6E1-

5 E3+ contains Ad5 sequences from bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). pAd6E1-E3+ contains the coding sequences for all Ad6 virion structural proteins which constitute its serotype specificity.

A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid 10 containing Ad6 regions is illustrated in Figure 11. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad6 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 35460 to 35759) and left (bp 1 to 450 and bp 3508 to 3807) end of the Ad6 genome separated 15 by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. These three segments were generated by PCR and cloned sequentially into pNEB193, generating pNEBAd6-3 (the ITR cassette). The ITR cassette contains a deletion of E1 sequences from Ad5 451 to 3507. The Ad6 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral 20 DNA in which recombination can occur.

Construction of pAd6 E1-E3- pre-adenovirus plasmids

Ad6 based vectors containing Ad5 regions and deleted in the E3 region were constructed starting with pAd6E1-E3+ containing Ad5 regions. A 5322 bp 25 subfragment of pAd6E1-E3+ containing the E3 region (Ad6 bp 25871 to 31192) was subcloned into pABS.3 generating pABSAd6E3. Three E3 deletions were then made in this plasmid generating three new plasmids pABSAd6E3(1.8Kb) (deleted for Ad6 bp 28602 to 30440), pABSAd6E3(2.3Kb) (deleted for Ad6 bp 28157 to 30437) and pABSAd6E3(2.6Kb) (deleted for Ad6 bp 28157 to 30788). Bacterial recombination 30 was then used to substitute the three E3 deletions back into pAd6E1-E3+ generating the Ad6 genome plasmids pAd6E1-E3-1.8Kb, pAd6E1-E3-2.3Kb and pAd6E1-E3-2.6Kb.

Example 6: Generation of Ad5 Genome Plasmid with the NS Sequence

A pcDNA3 plasmid (Invitrogen) containing the coding region NS3-NS4A-NS4B-NS5A was digested with *Xmn*I and *Nru*I restriction sites and the DNA fragment containing the CMV promoter, the NS3-NS4A-NS4B-NS5A coding sequence and the Bovine Growth Hormone (BGH) polyadenylation signal was cloned into the unique *Eco*RI restriction site of the shuttle vector pDelE1Spa, generating the Sva3-5A vector.

5 A pcDNA3 plasmid containing the coding region NS3-NS4A-NS4B-NS5A-NS5B was digested with *Xmn*I and *Eco*RI (partial digestion), and the DNA fragment containing part of NS5A, NS5B gene and the BGH polyadenylation signal was cloned into the Sva3-5A vector, digested *Eco*RI and *Bgl*II blunted with Klenow, generating the Sva3-5B vector.

10 The Sva3-5B vector was finally digested *Ssp*I and *Bst*1107I restriction sites and the DNA fragment containing the expression cassette (CMV promoter, NS3-NS4A-NS4B-NS5A-NS5B coding sequence and the BGH polyadenylation signal) flanked by adenovirus sequences was co-transformed with pAd5HVO (E1-,E3-) *Cla*I linearized genome plasmid into the bacterial strain BJ5183, to generate pAd5HVONS. 15 pAd5HVO contains Ad5 bp 1 to 341, bp 3525 to 28133 and bp 30818 to 35935.

Example 7: Generation of Adenovirus Genome Plasmids with the NSmut Sequence

20 Adenovirus genome plasmids containing an NS-mut sequence were generated in an Ad5 or Ad6 background. The Ad6 background contained Ad5 regions at bases 1 to 450, 3511 to 5548 and 33967 to 35935.

25 pV1JNS3-5Akozak was digested with *Bgl*II and *Xba*I restriction enzymes and the DNA fragment containing the Kozak sequence and the sequence coding NS3-NS4A-NS4B-NS5A was cloned into a *Bgl*II and *Xba*I digested polypMRKpdelE1 shuttle vector. The resulting vector was designated shNS3-5Akozak.

30 PolypMRKpdelE1 is a derivative of RKpdelE1(Pac/pIX/pack450) + CMVmin+BGH_pA(str.) modified by the insertion of a polylinker containing recognition sites for *Bgl*II, *Pme*I, *Swa*I, *Xba*I, *Sal*I, into the unique *Bgl*II restriction site present downstream the CMV promoter. MRKpdelE1(Pac/pIX/pack450) + CMVmin + BGH_pA(str.) contains Ad5 sequences from bp 1 to 5792 with a deletion of E1 sequences from bp 451 to 3510. The human CMV promoter and BGH polyadenylation signal were inserted into the E1 deletion in an E1 parallel orientation, with a unique *Bgl*II site separating them.

The NS5B fragment, mutated to abrogate enzymatic activity and with a strong translation termination at the 3' end, was obtained by assembly PCR and inserted into the shNS3-5Akozak vector via homologous recombination, generating polypMRKpdelE1NSmut. In polypMRKpdelE1NSmut the NS-mut coding sequence 5 is under the control of CMV promoter and the BGH polyadenylation signal is present downstream.

The gene expression cassette and the flanking regions which contain adenovirus sequences allowing homologous recombination were excised by digestion with *PacI* and *Bst*1107I restriction enzymes and co-transformed with either 10 pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb *Cla*I linearized genome plasmids into the bacterial strain BJ5183, to generate pAd5HVONSMut and pAd6E1-,E3-NSmut, respectively.

pAd6E1-E3-2.6Kb contains Ad5 bp 1 to 341 and from bp 3525 to 15 5548, Ad6 bp 5542 to 28157 and from bp 30788 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). In both plasmids the viral ITR's are joined by plasmid sequences that contain the bacterial origin of replication and an ampicillin resistance gene.

Example 8: Generation of Adenovirus Genome Plasmids with the NSOPTmut

The human codon-optimized synthetic gene (NSOPTmut) provided by 20 SEQ. ID. NO. 3 cloned into a pCRBlunt vector (Invitrogen) was digested with *Bam*H1 and *Sal*I restriction enzymes and cloned into *Bgl*II and *Sal*I restriction sites present in the shuttle vector polypMRKpdelE1. The resulting clone 25 (polypMRKpdelE1NSOPTmut) was digested with *Pac*I and *Bst*1107I restriction enzymes and co-transformed with either pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb *Cla*I linearized genome plasmids, into the bacterial strain BJ5183, to generate pAd5HVONSOPTmut and pAd6E1-,E3-NSOPTmut, respectively.

Example 9: Rescue and Amplification of Adenovirus Vectors

Adenovectors were rescued in Per.C6 cells. Per.C6 were grown in 10% 30 FCS / DMEM supplemented by L-glutamine (final 4mM), penicillin/streptomycin (final 100 IU/ml) and 10 mM MgCl₂. After infection, cells were kept in the same medium supplemented by 5% horse serum (HS). For viral rescue, 2.5 X 10⁶ Per.C6 were plated in 6 cm ø Petri dishes.

Twenty-four hours after plating, cells were transfected by calcium phosphate method with 10 µg of the *Pac I* linearized adenoviral DNA. The DNA precipitate was left on the cells for 4 hours. The medium was removed and 5% HS/DMEM was added.

5 Cells were kept in a CO₂ incubator until a cytopathic effect was visible (1 week). Cells and supernatant were recovered and subjected to 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). The lysate was centrifuged at 3000 rpm at - 4°C for 20 minutes and the recovered supernatant (corresponding to a cell lysate containing virus passed on cells only once; P1) was used, in the amount of 1 ml/ dish, 10 to infect 80-90% confluent Per.C6 in 10 cm ø Petri dishes. The infected cells were incubated until a cytopathic effect was visible, cells and supernatant recovered and the lysate prepared as described above (P2).

15 P2 lysate (4 ml) were used to infect 2 X 15 cm ø Petri dishes. The lysate recovered from this infection (P3) was kept in aliquots at -80°C as a stock of virus to be used as starting point for big viral preparations. In this case, 1 ml of the stock was enough to infect 2 X 15 cm ø Petri dishes and resulting lysate (P4) was used for the infection of the Petri dishes devoted to the large scale infection.

20 Further amplification was obtained from the P4 lysate which was diluted in medium without FCS and used to infect 30 X 15 cm ø Petri dishes (with Per.C6 80%-90% confluent) in the amount of 10 ml/dish. Cells were incubated 1 hour in the CO₂ incubator, mixing gently every 20 minutes. 12 ml / dish of 5% HS / DMEM was added and cells were incubated until a cytopathic effect was visible (about 48 hours).

25 Cells and supernatant were collected and centrifuged at 2K rpm for 20 minutes at 4°C. The pellet was resuspended in 15 ml of 0.1 M Tris pH=8.0. Cells were lysed by 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). 150 µl of 2 M MgCl₂ and 75 µl of DNase (10 mg of bovine pancreatic deoxyribonuclease I in 10 ml of 20 mM Tris-HCl pH= 7.4, 50 mM NaCl, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, 50% glycerol) were added. After a 1 hour incubation at 37°C 30 in a water bath (vortex every 15 minutes) the lysate was centrifuged at 4K rpm for 15 minutes at 4°C. The recovered supernatant was ready to be applied on CsCl gradient.

35 The CsCl gradients were prepared in SW40 ultra-clear tubes as follows:
0.5 ml of 1.5d CsCl
3 ml of 1.35d CsCl

3 ml of 1.25d CsCl

5-ml/ tube of viral supernatant was applied.

If necessary, the tubes were topped up with 0.1 M tris-Cl pH=8.0.

Tubes were centrifuged at 35K rpm for 1 hour at -10°C with rotor SW40. The viral bands (located at the 1.25/1.35 interface) were collected using a syringe.

The virus was transferred into a new SW40 ultraclear tube and 1.35d CsCl was added to top the tube up. After centrifugation at 35K rpm for 24 hours at 10°C in the rotor SW40, the virus was collected in the smallest possible volume and dialyzed extensively against buffer A105 (5 mM Tris, 5% sucrose, 75 mM NaCl, 1 mM MgCl₂, 0.005% polysorbate 80 pH=8.0). After dialysis, glycerol was added to final 10% and the virus was stored in aliquots at -80°C.

Example 10: Enhanced Adenovector Rescue

First generation Ad5 and Ad6 vectors carrying HCV NSOPTmut transgene were found to be difficult to rescue. A possible block in the rescue process might be attributed to an inefficient replication of plasmid DNA that is a sub-optimal template for the replication machinery of adenovirus. The absence of the terminal protein linked to the 5'ends of the DNA (normally present in the viral DNA), associated with the very high G-C content of the transgene inserted in the E1 region of the vector, may be causing a substantial reduction in replication rate of the plasmid-derived adenovirus.

To set up a more efficient and reproducible procedure for rescuing Ad vectors, an expression vector (pE2; Figure 19) containing all E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 under the control of tet-inducible promoter was employed. The transfection of pE2 in combination with a normal preadeno plasmid in PerC6 and in 293 leads to a strong increase of Ad DNA replication and to a more efficient production of complete infectious adenovirus particles.

30. Plasmid Construction

pE2 is based on the cloning vector pBI (CLONTECH) with the addition of two elements to allow episomal replication and selection in cell culture: (1) the EBV-OriP (EBV [nt] 7421-8042) region permitting plasmid replication in synchrony with the cell cycle when EBNA-1 is expressed and (2) the hygromycin-B phosphotransferase (HPH)-resistance gene allowing a positive selection of

transformed cells. The two transcriptional units for the adenoviral genes E2 a and b and E4-Orf6 were constructed and assembled in pE2 as described below.

5 The Ad5-Polymerase *Clal/SphI* fragment and the Ad5-pTP *Acc65/EcoRV* fragment were obtained from pVac-Pol and pVac-pTP (Stunnenberg *et al.* *NAR* 16:2431-2444, 1988). Both fragments were filled with Klenow and cloned into the *Sall* (filled) and *EcoRV* sites of pBI, respectively obtaining pBI-Pol/pTP.

10 EBV-OriP element from pCEP4 (Invitrogen) was first inserted within two chicken β -globin insulator dimers by cloning it into *BamHI* site of pJC13-1 (Chung *et al.*, *Cell* 74(3):505-14, 1993). HS4-OriP fragment from pJC13-OriP was then cloned inside pSA1mv (a plasmid containing tk-Hygro-B resistance gene expression cassette as well as Ad5 replication origin), the ITR's arranged as head-to-tail junction, obtained by PCR from pFG140 (Graham, *EMBO J.* 3:2917-2922, 1984) using the following primers: 5'-TCGAATCGATCGCGAACCTACGC-3' (SEQ. ID. NO. 16) and 5'-TCGACGTGTCGACTTCGAAGCGCACACCAAAAAACGTC-3' (SEQ. ID. NO. 17), thus generating pMVHS4Orip. A DNA fragment from pMVHS4Orip, containing the insulated OriP, Ad5 ITR junction and tk-HygroB cassette, was then inserted into pBI-Pol/pTP vector restricted *AseII/AatII* generating pBI-Pol/pTPHS4.

20 To construct the second transcriptional unit expressing Ad5-Orf6 as well as Ad5-DBP, E4orf6 (Ad 5 [nt] 33193-34077) obtained by PCR was first inserted into pBI vector, generating pBI-Orf6. Subsequently, DBP coding DNA sequence (Ad 5 [nt] 22443-24032) was inserted into pBI-Orf6 obtaining the second bi-directional Tet-regulated expression vector (pBI-DBP/E4orf6). The original polyA signals present in pBI were substituted with BGH and SV40 polyA.

25 pBI-DBP/E4orf6 was then modified by inserting a DNA fragment containing the Adeno5-ITRs arranged in head-to-tail junction plus the hygromycin B resistance gene obtained from plasmid pSA-1mv. The new plasmid pBI-DBP/E4orf6shuttle was then used as donor plasmid to insert the second tet-regulated transcriptional unit into pBI-Pol/pTPHS4 by homologous recombination using *E. coli* strain BJ5183 obtaining pE2.

Cell lines, Transfections and Virus Amplification

30 PerC6 cells were cultured in Dulbecco's modified Eagle's Medium (DMEM) plus 10% fetal bovine serum (FBS), 10 mM MgCl₂, penicillin (100 U/ml), streptomycin (100 μ g/ml) and 2 mM glutamine.

All transient transfections were performed using Lipofectamine2000 (Invitrogen) as described by the manufacturer. 90% confluent PERC.6™ planted in 6-cm plates were transfected with 3.5 µg of Ad5/6NSOPTmut pre-adeno plasmids, digested with PacI, alone or in combination with 5 µg pE2 plus 1 µg pUHD52.1.

5 pUHD52.1 is the expression vector for the reverse tet transactivator 2 (rtTA2) (Urlinger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 97(14):7963-7968, 2000). Upon transfection, cells were cultivated in the presence of 1 µg/ml of doxycycline to activate pE2 expression. 7 days post-transfection cells were harvested and cell lysate was obtained by three cycles of freeze-thaw. Two ml of cell lysate were used to infect

10 a second 6-cm dish of PerC6. Infected cells were cultivated until a full CPE was observed then harvested. The virus was serially passaged five times as described above, then purified on CsCl gradient. The DNA structure of the purified virus was controlled by endonuclease digestion and agarose gel electrophoresis analysis and compared to the original pre-adeno plasmid restriction pattern.

15

Example 11: Partial Optimization of HCV Polyprotein Encoding Nucleic acid

Partial optimization of HCV polyprotein encoding nucleic acid was performed to facilitate the production of adenovectors containing codons optimized for expression in a human host. The overall objective was to provide for increased expression due to codon optimization, while facilitating the production of an adenovector encoding HCV polyprotein.

20

Several difficulties were encountered in producing an adenovector encoding HCV polyprotein with codons optimized for expression in a human host. An adenovector containing an optimized sequence (SEQ. ID. NO. 3) was found to be more difficult to synthesize and rescue than an adenovector containing a non-optimized sequence (SEQ. ID. NO. 2).

25

The difficulties in producing an adenovector containing SEQ. ID. NO. 3 were attributed to a high GC content. A particularly problematic region was the region at about position 3900 of NSOPTmut (SEQ. ID. NO. 3).

30 Alternative versions of optimized HCV encoding nucleic acid sequence were designed to facilitate its use in an adenovector. The alternative versions, compared to NSOPTmut, were designed to have a lower overall GC content, to reduce/avoid the presence of potentially problematic motifs of consecutive G's or C's, while maintaining a high level of codon optimization to allow improved expression of the encoded polyprotein and the individual cleavage products.

35

A starting point for the generation of a suboptimally codon-optimized sequence is the coding region of the NSOPTmut nucleotide sequence (bases 7 to 5961 of SEQ. ID. NO. 3). Values for codon usage frequencies (normalized to a total of 1.0 for each amino acid) were taken from the file human_high.cod available in the 5 Wisconsin Package Version 10.3 (Accelrys Inc., a wholly owned subsidiary of Pharmacopeia, Inc.).

To reduce the local and overall GC content a table defining preferred codon substitutions for each amino acid was manually generated. For each amino acid the codon having 1) a lower GC content as compared to the most frequent codon and 10 2) a relatively high observed codon usage frequency (as defined in human_high.cod) was chosen as the replacement codon. For example for Arg the codon with the highest frequency is CGC. Out of the other five alternative codons encoding Arg (CGG, AGG, AGA, CGT, CGA) three (AGG, CGT, CGA) reduce the GC content by 1 base, one (AGA) by two bases and one (CGG) by 0 bases. Since the AGA codon is 15 listed in human_high.cod as having a relatively low usage frequency (0.1), the codon substituting CGC was therefore chosen to be AGG with a relative frequency of 0.18. Similar criteria were applied in order to establish codon replacements for the other amino acids resulting in the list shown in Table 5. Parameters applied in the following optimization procedure were determined empirically such that the resulting sequence 20 maintained a considerably improved codon usage (for each amino acid) and the GC content (overall and in form of local stretches of consecutive G's and/or C's) was decreased.

Two examples of partial optimized HCV encoding sequences are provided by SEQ. ID. NO. 10 and SEQ. ID. NO. 11. SEQ. ID. NO. 10 provides a 25 HCV encoding sequence that is partially optimized throughout. SEQ. ID. NO. 11 provides an HCV encoding sequence fully optimized for codon usage with the exception of a region that was partially optimized.

Codon optimization was performed using the following procedure:
Step 1) The coding region of the input fully optimized NSOPTmut 30 sequence was analyzed using a sliding window of 3 codons (9 bases) shifting the window by one codon after each cycle. Whenever a stretch containing 5 or more consecutive C's and/or G's was detected in the window the following replacement rule was applied: Let N indicate the number of codon replacements previously performed. If N is odd replace the middle codon in the window with the codon specified in Table 35 5, if N is even replace the third terminal codon in the window with the codon

specified in a codon optimization table such as human_high.cod. If Leu or Val is present at the second or third codon do not apply any replacement in order not to introduce Leu or Val codons with very low relative codon usage frequency (see, for example, human_high.cod). In the following cycle analysis of the shifted window was

5 then applied to a sequence containing the replacements of the previous cycle.

The alternating replacement of the middle and terminal codon in the 3 codon window was found empirically to give a more satisfying overall maintenance of optimized codon usage while also reducing GC content (as judged from the final sequence after the procedure). In general, however, the precise replacement strategy

10 depends on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

Step 2) The sequence containing all the codon replacements performed during step 1) was then subjected to an additional analysis using a sliding window of 21 codons (63 bases) in length: according to an adjustable parameter the overall GC

15 content in the window was determined. If the GC content in the window was higher than 70% the following codon replacement strategy was applied: In the window replace the codons for the amino acids Asn, Asp, Cys, Glu, His, Ile, Lys, Phe, Tyr by the codons given in Table 5. Restriction of the replacement to this set of amino acids was motivated by the fact that a) the replacement codon still has an acceptably high

20 frequency of usage in human_high.cod and b) the average overall human codon usage in CUTG for the replacement codon is nearly as high as the most frequent codon. In the following cycle analysis of the shifted window is then applied to a sequence containing the replacements of the previous cycle.

The threshold 70% was determined empirically by compromising between an overall reduction in GC content and maintenance of a high codon optimization for the individual amino acids. As in step 1) the precise replacement strategy (choice of amino acids and GC content threshold value) will again depend on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

30 Step 3) The sequence generated by steps 1) and 2) was then manually edited and additional codons were changed according to the following criteria: Regions still having a GC content higher than 70% over a window of 21 codons were examined manually and a few codons were replaced again following the scheme given in Table 5.

Subsequent steps were performed to provide for useful restriction sites, remove possible open reading frames on the complementary strand, to add homologous recombinant regions, to add a Kozac signal, and to add a terminator.

These steps are numbered 4-7

5 Step 4) The sequence generated in step 3 was examined for the absence of certain restriction sites (BglII, PmeI and XbaI) and presence of only 1 StuI site to allow a subsequent cloning strategy using a subset of restriction enzymes. Two sites (one for BglII and one for StuI) were removed from the sequence by replacing codons that were part of the respective recognition sites.

10 Step 5) The sequence generated by steps 1) through 4) was then modified according to allow subsequent generation of a modified NSOPTmut sequence (by homologous recombination). In the sequence obtained from steps 1) through 4) the segment comprising base 3556 to 3755 and the segment comprising base 4456 to 4656 were replaced by the corresponding segments from NSOPTmut.

15 15 The segment comprising bases 3556 to 4656 of SEQ. ID. NO. 10 can be used to replace the problematic region in NSOPTmut (around position 3900) by homologous recombination thus creating the variant of NSOPTmut having the sequence of SEQ. ID. NO. 11.

20 Step 6) Analysis of the sequence generated through steps 1) to 5) revealed a potential open reading frame spanning nearly the complete fragment on the complementary strand. Removal of all codons CTA and TTA (Leu) and TCA (Ser) from the sense strand effectively removed all stop codons in one of the reading frames on the complementary strand open reading frame and subsequent translation into protein is very small, in order to exclude a potential interference with the transcription and subsequent translation of the sequence encoded on the sense strand, TCA codons for Ser were introduced on the sense approximately every 500 bases. No changes were introduced in the segments introduced during step 5) to allow homologous recombination. The TCA codon for Ser was preferred over the CTA and TTA codons for Leu because of the higher relative frequency for TCA (0.05) as compared to CTA (0.02) and TTA (0.03) in human_high.cod. In addition, the average human codon usage from CUTG favored TCA (0.14 against 0.07 for CTA and TTA).

25 30 35 Step 7) In a final step GCCACC was added at the 5' end of the sequence to generate an optimized internal ribosome entry site (Kozac signal) and a TAAA stop signal was added at the 3'. To maintain the initiation of translation

properties of NSsuboptmut the first 8 codons of the coding region were kept identical to the NSOPTmut sequence. The resulting sequence was again checked for the absence of *Bgl*II, *Pme*I and *Xba*I recognition sites and the presence of only 1 *Stu*I site.

The NSsuboptmut sequence (SEQ. ID. NO. 10) has an overall reduced 5 GC content (63.5%) as compared to NSOPTmut (70.3%) and maintains a well optimized level of codon usage optimization. Nucleotide sequence identity of NSsuboptmut is 77.2% with respect to NSmut.

Table 5: Definition of codon replacements performed during steps 1) and 2).

Amino Acid	Most frequent codon	Relative frequency	Reduction in GC content (bases)	Replacement codon	Relative frequency
Amino Acids where the replacement codon reduces the codon GC-content by 1 base					
Ala	GCC	0.51	1	GCT	0.17
Arg	CGC	0.37	1	AGG	0.18
Asn	AAC	0.78	1	AAT	0.22
Asp	GAC	0.75	1	GAT	0.25
Cys	TGC	0.68	1	TGT	0.32
Glu	GAG	0.75	1	GAA	0.25
Gln	CAG	0.88	1	CAA	0.12
Gly	GGC	0.50	1	GGA	0.14
His	CAC	0.79	1	CAT	0.21
Ile	ATC	0.77	1	ATT	0.18
Lys	AAG	0.82	1	AAA	0.18
Phe	TTC	0.80	1	TTT	0.20
Pro	CCC	0.48	1	CCT	0.19
Ser	AGC	0.34	1	TCT	0.13
Thr	ACC	0.51	1	ACA	0.14
Tyr	TAC	0.74	1	TAT	0.26
Amino Acids with no alternative codon					
Met	ATG	1.00	0	ATG	1.00
Trp	TGG	1.00	0	TGG	1.00

Amino Acids where the replacement codon has a very low relative frequency. These amino acids were excluded from the replacement procedure					
Leu	CTG	0.58	1	TTG	0.06
Val	GTG	0.64	1	GTT	0.07

Example 12: Virus Characterization

5 Adenovectors were characterized by: (a) measuring the physical particles/ml; (b) running a TaqMan PCR assay; and (c) checking protein expression after infection of HeLa cells.

10 a) *Physical Particles Determination*

15 CsCl purified virus was diluted 1/10 and 1/100 in 0.1% SDS PBS. As a control, buffer A105 was used. These dilutions were incubated 10 minutes at 55°C. After spinning the tubes briefly, O.D. at 260 nm was measured. The amount of viral particles was calculated as follows: 1 OD 260 nm = 1.1×10^{12} physical particles/ml. The results were typically between 5×10^{11} and 1×10^{12} physical particles /ml.

20 b) *TaqMan PCR Assay*

25 TaqMan PCR assay was used for adenovectors genome quantification (Q-PCR particles/ml). TaqMan PCR assay was performed using the ABI Prism 7700 sequence detector. The reaction was performed in a final 50 μ l volume in the presence of oligonucleotides (at final 200 nM) and probe (at final 200 μ M) specific for the adenoviral backbone. The virus was diluted 1/10 in 0.1% SDS PBS and incubated 10 minutes at 55°C. After spinning the tube briefly, serial 1/10 dilutions (in water) were prepared. 10 μ l the 10^{-3} , 10^{-5} and 10^{-7} dilutions were used as templates in the PCR assay.

25 The amount of particles present in each sample was calculated on the basis of a standard curve run in the same experiment. Typically results were between 1×10^{12} and 3×10^{12} Q-PCR particles /ml.

30 c) *Expression of HCV Non-Structural Proteins*
Expression of HCV NS proteins was tested by infection of HeLa cells. Cells were plated the day before the infection at 1.5×10^6 cells/dish (10 cm \varnothing Petri dishes). Different amounts of CsCl purified virus corresponding to m.o.i. of 50, 250

and 1250 pp/cell were diluted in medium (FCS free) up to a final volume of 5 ml. The diluted virus was added on the cells and incubated for 1 hour at 37°C in a CO₂ incubator (gently mixing every 20 minutes). 5 ml of 5% HS-DMEM was added and the cells were incubated at 37°C for 48 hours.

5 Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were run on 10% SDS-acrylamide gel, blotted on nitrocellulose and assayed with antibodies directed against NS3, NS5a and NS5b in order to check the correct polyprotein cleavage. Mock-infected cells were used as a negative control. Results from representative experiments testing the Ad5-NS, MRKAd5-NSmut, MRKAd6-
10 NSmut and MRKAd6-NSOPTmut are shown in Figure 14.

Example 13: Mice Immunization with Adenovectors Encoding Different NS Cassettes

15 The adenovectors Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut and MRKAd6-NSOPTmut were injected in C57Black6 mice strains to evaluate their potential to elicit anti-HCV immune responses. Groups of animals (N=9-10) were injected intramuscularly with 10⁹ pp of CsCl purified virus. Each animal received two doses at three weeks interval.

20 Humoral immune response against the NS3 protein was measured in post dose two sera from C57Black6 immunized mice by ELISA on bacterially expressed NS3 protease domain. Antibodies specific for the tested antigen were detected with geometric mean titers (GMT) ranging from 100 to 46000 (Tables 6, 7, 8 and 9).

25

Table 6: Ad5-NS

Mice n.	1	2	3	4	5	6	7	8	9	10	GMT
Titer	50	253	50	50	50	2257	504	50	50	50	108

30

Table 7: Ad5-NSmut

Mice n.	11	12	13	14	15	16	17	18	19	20	GMT
Titer	3162	78850	87241	6796	12134	3340	18473	13093	76167	49593	23645

Table 8: MRKAd6-NSmut

Mice n.	21	22	23	24	25	26	27	28	29	30	GMT
Titer	125626	39751	40187	65834	60619	69933	21555	49348	29290	26859	46461

Table 9: MRKAd6-NSOPTmut

Mice n.	31	32	33	34	35	36	37	GMT
Titer	25430	3657	893	175	10442	49540	173	2785

10 T cell response in C57Black6 mice was analyzed by the quantitative
 15 ELISPOT assay measuring the number of IFN γ secreting T cells in response to five pools (named from F to L+M) of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response induced in C57Black6 mice was analyzed by the same assay using a 20mer peptide encompassing a CD8+ epitope for C57Black6 mice (pep1480). Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELIspot assay.

20 Spleen cells, splenocytes and peptides were produced and treated as described in Example 3, *supra*. Representative data from groups of C57Black6 mice (N=9-10) immunized with two injections of 10^9 viral particles of vectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are shown in Figure 15.

Example 14: Immunization of Rhesus macaques with Adenovectors

Rhesus macaques (N=3-4) were immunized by intramuscular injection of CsCl purified Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut or MRKAd6-

NSOPTmut virus. Each animal received two doses of 10^{11} or 10^{10} vp in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by a) IFN- γ ELISPOT (see Example 3, *supra*), b) IFN- γ ICS and c) bulk CTL assays. These assays measure HCV 5 antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid 10 sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5a, NS5b) were prepared for use in these assays to measure immune responses in HCV DNA and adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large 15 pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

IFN- γ ICS:

For IFN- γ ICS, 2×10^6 PBMC in 1 ml R10 (RPMI medium, 20 supplemented with 10% FCS) were stimulated with peptide pool antigens. Final concentration of each peptide was 2 μ g/ml. Cells were incubated for 1 hour in a CO₂ incubator at 37°C and then Brefeldin A was added to a final concentration of 10 μ g /ml to inhibit the secretion of soluble cytokines. Cells were incubated for additional 14-16 hours at 37°C.

25 Stimulation was done in the presence of co-stimulatory antibodies: CD28 and CD49d (anti-humanCD28 BD340975 and anti-humanCD49d BD340976). After incubation, cells were stained with fluorochrome-conjugated antibodies for surface antigens: anti-CD3, anti-CD4, anti-CD8 (CD3-APC Biosource APS0301, CD4-PE BD345769, CD8-PerCP BD345774).

30 To detect intracellular cytokines, cells were treated with FACS permeabilization buffer 2 (BD340973), 2x final concentration. Once fixed and permeabilized, cells were incubated with an antibody against human IFN- γ , IFN- γ FITC (Biosource AHC4338).

Cells were resuspended in 1% formaldehyde in PBS and analyzed at 35 FACS within 24 hours. Four color FACS analysis was performed on a FACSCalibur

instrument (Becton Dickinson) equipped with two lasers. Acquisition was done gating on the lymphocyte population in the Forward versus Side Scatter plot coupled with the CD3, CD8 positive populations. At least 30,000 events of the gate were taken. The positive cells are expressed as number of IFN- γ expressing cells over 10^6 lymphocytes.

IFN- γ ELISPOT and IFN- γ ICS data from immunized monkeys after one or two injections of 10^{10} or 10^{11} vp of the different adenovectors are reported in Figures 16A-16D, 17A, and 17B.

10 **Bulk CTL Assays**

A distinguishing effector function of T lymphocytes is the ability of subsets of this cell population to directly lyse cells exhibiting appropriate MHC-associated antigenic peptides. This cytotoxic activity is most often associated with CD8+ T lymphocytes.

15 PBMC samples were infected with recombinant vaccine viruses expressing HCV antigens *in vitro* for approximately 14 days to provide antigen restimulation and expansion of memory T cells. Cytotoxicity against autologous B cell lines treated with peptide antigen pools was tested.

20 The lytic function of the culture is measured as a percentage of specific lysis resulted from chromium released from target cells during 4 hours incubation with CTL effector cells. Specific cytotoxicity is measured and compared to irrelevant antigen or excipient-treated B cell lines. This assay is semi-quantitative and is the preferred means for determining whether CTL responses were elicited by the vaccine. Data after two injections from monkeys immunized with 10^{11} vp/dose with 25 adenovectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are reported in Figures 18A-18F.

30 Other embodiments are within the following claims. While several embodiments have been shown and described, various modifications may be made without departing from the spirit and scope of the present invention.

WHAT IS CLAIMED IS:

1. A nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1, provided that said polypeptide has sufficient protease activity to process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive.
2. The nucleic acid of claim 1, wherein said nucleotide sequence is substantially similar to the coding sequence of SEQ ID NO: 2.
3. The nucleic acid of claim 1, wherein said nucleotide sequence encodes for the polypeptide of SEQ ID NO: 1.
4. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
5. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2 or SEQ ID NO: 3.
6. The nucleic acid of any one of claims 1-5, wherein said nucleic acid is an expression vector capable of expressing said polypeptide from said nucleotide sequence in a human cell.
7. A nucleic acid comprising a gene expression cassette able to express a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1 in a human cell, provided that said polypeptide can process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive, said expression cassette comprising:
 - a) a promoter transcriptionally coupled to a nucleotide sequence encoding said polypeptide;
 - b) a 5' ribosome binding site functionally coupled to said nucleotide sequence,

c) a terminator joined to the 3' end of said nucleotide sequence, and
d) a 3' polyadenylation signal functionally coupled to said nucleotide sequence.

5 8. The nucleic acid of claim 7, wherein said nucleotide sequence is substantially similar to either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

10 9. The nucleic acid of claim 8, wherein said nucleic acid is a shuttle vector further comprising a selectable marker, an origin of replication, a first adenovirus homology region and a second adenovirus homology region flanking said expression cassette, wherein said first homology region has at least about 100 base pairs substantially homologous to at least right end of a wild-type adenovirus region from about base pairs 1-425, and said second homology region has at least about 100 base pairs substantially homologous to at least the left end of a wild-type adenovirus region from about base pairs 3511-5792 of Ad5 or corresponding region of another adenovirus.

20 10. The nucleic acid of claim 9, wherein said nucleotide sequence encodes for a polypeptide of SEQ ID NO: 1.

11. The nucleic acid of claim 9, wherein said nucleotide sequence is SEQ ID NO: 2.

25 12. The nucleic acid of claim 9, wherein said nucleotide sequence is either SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

30 13. The nucleic acid of claim 8, wherein said nucleic acid is a plasmid suitable for administration into a human and further comprises a prokaryotic origin of replication and a gene coding for a selectable marker.

14. The nucleic acid of claim 13, wherein said nucleotide sequence encodes for a polypeptide of SEQ ID NO: 1.

15. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

5 16. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of SEQ ID NO: 2 or SEQ ID NO: 3.

10 17. The nucleic acid of claim 14, wherein said promoter is the human intermediate early cytomegalovirus promoter (intron A), said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the bovine growth hormone (BGH) polyadenylation signal.

15 18. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising a selectable marker, an origin of replication, and a recombinant adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette.

20 19. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising a selectable marker, an origin of replication, and

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

5 20. The nucleic acid of claim 19, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

10 21. The nucleic acid of claim 20, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

15 22. The nucleic acid of claim 21, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

20 23. The nucleic acid of claim 19, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

25 24. The nucleic acid of claim 23, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

30 25. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

35 26. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2 or SEQ ID NO: 3.

27. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising an origin of replication, a selectable marker, and:

- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- 10 c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- 15 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base 20 pair 35759 corresponding to Ad6, joined to said fourth region.

25 28. The nucleic acid of claim 27, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

30 29. The nucleic acid of claim 28, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

35 30. The nucleic acid of claim 27, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

31. The nucleic acid of claim 30, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

32. The nucleic acid of claim 8, wherein said nucleic acid is a adenovector consisting of a nucleotide sequence substantially similar to of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV 10 polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.

33. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector having an adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette

34. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:

- 20 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
- c) a second adenovirus region from about base pair 3511 to about 25 base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- 30 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
- f) a fifth adenovirus region from about base pair 33967 to about 35 base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

35. The nucleic acid of claim 34, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

5

36. The nucleic acid of claim 35, wherein said promoter is the

human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

10

37. The nucleic acid of claim 36, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

15

38. The nucleic acid of claim 34, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

20

39. The nucleic acid of claim 37, where said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

25

40. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

30

41. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2 or SEQ ID NO: 3.

42. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

43. The nucleic acid of claim 42, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

44. The nucleic acid of claim 42, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

45. An adenovector consisting of the nucleic acid sequence of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.

46. An adenovector produced by a process comprising the steps of:

- a) producing an adenovirus genome plasmid by homologous recombination between the shuttle vector of claim 9 and a nucleic acid comprising:
 - a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - 5 a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
 - a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 10 28156 corresponding to Ad6, joined to said second region;
 - a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
 - 15 a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region; and
- b) rescuing said adenovector from said adenovirus plasmid.

47. A cultured recombinant cell comprising the nucleic acid of 20 claim 6.

48. A cultured recombinant cell comprising the nucleic acid of any one of claims 9-46.

25 49. A method of making an adenovector comprising the steps of:

- a) producing an adenovirus genome plasmid comprising a gene expression cassette by homologous recombination between the nucleic acid of claim 9 and a nucleic acid comprising:
 - a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - 30 a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

15 a third adenovirus region from about base pair 5549 to about
base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair
28156 corresponding to Ad6, joined to said second region;
a fourth adenovirus region from about base pair 30818 to about
5 base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base
pair 33784 corresponding to Ad6, joined to said third region; and
a fifth adenovirus region from about base pair 33967 to about
base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base
pair 35759 corresponding to Ad6, joined to the fourth region; and
10 b) rescuing said recombinant adenovirus from said recombinant
adenovirus plasmid.

20 50. A pharmaceutical composition comprising the nucleic acid of
any one of claims 13-17 and 32-46 and pharmaceutically acceptable carrier.

25 51. A method of treating a patient comprising the step of
administering to said patient an effective amount of the nucleic acid of any one of
claims 13-17 and 32-46.

30 52. The method of claim 51, wherein said patient is a human.

53. The method of claim 52, wherein said patient is not infected
with HCV.

54. The method of claim 52, wherein said patient is infected with
HCV.

35 55. A recombinant nucleic acid comprising one or more Ad6
regions and a region not present in Ad6, wherein at least one Ad6 region is selected
from the group consisting of: E1A, E1B, E2B, E2A, E4, L1, L2, L4, and L5.

56. The recombinant nucleic acid of claim 55, wherein said region
not present in Ad6, is an expression cassette coding for a polypeptide not found in
Ad6.

57. The recombinant nucleic acid of claim 56, wherein said recombinant nucleic acid is an adenovirus vector defective in at least E1 that is able to replicate when E1 is supplied *in trans*.

5 58. The recombinant nucleic acid of claim 57, wherein said vector
consists of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) said gene expression cassette in an E1 parallel or E1 anti-parallel orientation joined to said first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said gene expression cassette;

15 base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;

e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about 30789 corresponding to Ad6, joined to said third region;

20 f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, wherein said fifth region is joined to said fourth region if said fourth region is present, or said fifth is joined to said third region if said fourth region is not present; and

25 g) a sixth adenovirus region from about base pair 33967 to about
base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base
pair 35759 corresponding to Ad6, joined to said fourth region;
provided that at least one of said second, third, and fifth regions is
from Ad6.

30 59. The recombinant nucleic acid of claim 57, wherein said vector
consists of:

a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- 5 c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- 10 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;

15 provided that at least one of said second, third, and fourth regions is from Ad6.

1 MAPITAYSQQ TRGLLGCIIT SLTGRDKNQV EGEVQVVSTA TQSFLATCVN
51 GVCWTVYHGA GSCTLGPKG PITQMYTNVD QDLVWQAPP GARSLTPCTC
101 GSSDLYLVTR HADVIPVRRR GDSRGSSLSP RPVSYLKGSS GGPLLCPGSH
151 AVGIPRAAVC TRGVAKAVDF VPVESMETTM RSPVFTDNSS PPAVPQSFQV
201 AHLHAPTGSG KSTKVPAAAYA AQGYKVLVNL PSVAATLGFG AYMSKAHGID
251 PNIRTGVRTI TTGAPVTYST YGKFLADGGC SGGAYDIIIC DECHSTDSTT
301 ILGIGTVLDQ AETAGARLUVV LATATPPGSV TVPHPNIEEV ALSNTGEIPF
351 YGKAIPIEAI RGGRHLIFCH SKKKCDELAA KLSGLGINAV AYYRGLDVSV
401 IPTIGDVVVV ATDALMTGYT GDFDSVIDCN TCVTQTVDFSL LDPTFTIETT
451 TVPQDAVSRS QRRGRTGRGR RGIYRFVTPG ERPSGMFDSS VLCECYDAGC
501 AWYELTPAET SVRLRAYLNT PGLPVCQDHL EFWESVFTGL THIDAHFLSQ
551 TKQAGDNFPY LVAYQATVCA RAQAPPPSWD QMWKCLIRLK PTLHGPTPLL
601 YRLGAVQNEV TLTHPITKYI MACMSADLEV VTSTWVLVGG VLAALAAYCL
651 TTGSVVIVGR IILSGRPAIV PDREFLYQEF DEMEECASHL PYIEQGMQLA
701 EQFKQKALGL LQTATKQAEA AAPVVESKWR ALETFWAKHM WNFISGIQYL
751 AGLSTLPGNP AIASLMAFTA SITSPLTTQS TLLFNILGGW VAAQLAPPSA
801 ASAFTVGAGIA GAAVGSIGLG KVLVDILAGY GAGVAGALVA FKVMSGEMPS
851 TEDLVNLLPA ILSPGALVVG VVCAAILRRH VGPGEHAVQW MNRLIAFASR
901 GNHVSPTHYV PESDAAARVT QILSSLTITQ LLKRLHQWIN EDCSTPCSGS
951 WLRDVWDWIC TVLTDFTKTL QSKLLPQLPG VPFSCQRGY KGVWRGDGIM
1001 QTTCPGQAQI TGHVKNGSMR IVGPKTCNSNT WHGTFFPINAY TTGPCTPSPA
1051 PNYSRALWRV AAEYEVTR VGDFHYVTGM TTDNVKCPHQ VPAPEFFTEV
1101 DGVRLHRYAP ACRPLLREEV TFQVGLNQYL VGSQLPCEPE PDVAVLTSML
1151 TDPSHITAET AKRRLARGSP PSLASSSASQ LSAPSLKATC TTHHVSPDAD
1201 LIEANLLWRQ EMGGNITRVE SENKVVVLDs FDPLRAEDE REVSPAEIL
1251 RKSKKFPAAM PIWARPDYNP PLLESWKDPD YVPPVVHGCP LPIPIKAPP
1301 PPRRKRTVVL TESSVSSALA ELATKTFGSS ESSAVDSGTA TALPDQASDD
1351 GDKGSDVESY SSMPPLEGEV GDPDLSDGSW STVSEEASED VVCCSMSYTW
1401 TGALITPCAA EESKLPINAL SNSLLRHNM VYATTSRSAG LRQKKVTFDR
1451 LQVLDDHYRD VLKEMKAKAS TVKAKLLSVE EACKLTPPHS AKSKFGYGA
1501 DVRNLSSKAV NHIHSVWKDL LEDTVTPIDT TIMAKNEVFC VQPEKGGRK
1551 ARLIVFPDLG VRVCEKMALY DVVSTLPQVV MGSSYGFQYS PGQRVEFLVN
1601 TWKSKKNPMG FSYDTRCFDS TVTENDIRVE ESIYQCCDLA PEARQAIKSL
1651 TERLYIGGPL TNSKGQNCGY RR CRASGVLT TSCGNTLTCY LKASAACRAA

2/92

1701 KLQDCTMLVN AAGLVVICES AGTQEDAASL RVFTEAMTRY SAPPGDPPQP
1751 EYDLELITSC SSNVSAHDA SGKRVYYLTR DPTTPLARAA WETARHTPVN
1801 SWLGNIIIMYA PTLWARMILM THFFSILLAQ EQLEKALDCQ IYGACYSIEP
1851 LDLPQIIERL HGLSAFSLHS YSPGEINRVA SCLRKLGVPP LRVWRHRARS
1901 VRARLLSQGG RAATCGKYL F NWAVTKLKL TPIPAASQLD LSGWFVAGYS
1951 GGDIIYHSLSR ARPRWFMCL LLLSVGVGIVY LLPNR

FIG. 1B

3/92

1 GCCACCATGG CGCCCCATCAC GGCCTACTCC CAACAGACGC GGGGCCTACT
 51 TGGTTGCATC ATCACTAGCC TTACAGGCCG GGACAAGAAC CAGGTCGAGG
 101 GAGAGGTTCA GGTGGTTTCC ACCGCAACAC AATCCTTCCT GGCGACCTGC
 151 GTCAACGGCG TGTGTTGGAC CGTTTACCAT GGTGCTGGCT CAAAGACCTT
 201 AGCCGGCCCCA AAGGGGCCAA TCACCCAGAT GTACACTAAT GTGGACCAGG
 251 ACCTCGTCGG CTGGCAGGCG CCCCCCGGGG CGCGTTCCCTT GACACCATGC
 301 ACCTGTGGCA GCTCAGACCT TTACTTGGTC ACGAGACATG CTGACGTCA
 351 TCCGGTGCCTC CGGGGGGGCG ACAGTAGGGG GAGCCTGCTC TCCCCCAGGC
 401 CTGTCTCCTA CTTGAAGGGC TCTTCGGGTG GTCCACTGCT CTGCCCTTCG
 451 GGGCACGCTG TGGGCATCTT CCGGGCTGCC GTATGCACCC GGGGGGTTGC
 501 GAAGGCGGTG GACTTTGTGC CCGTAGAGTC CATGGAAACT ACTATGCGGT
 551 CTCCGGTCTT CACGACAAC TCATCCCCC CGGCCGTACC GCAGTCATTT
 601 CAAGTGGCCC ACCTACACGC TCCCACCTGGC AGCGGCAAGA GTACTAAAGT
 651 GCCGGCTGCA TATGCAGCCC AAGGGTACAA GGTGCTGCTC CTCAATCCGT
 701 CCGTTGCCGC TACCTTAGGG TTTGGGGCGT ATATGTCTAA GGCACACGGT
 751 ATTGACCCCCA ACATCAGAAC TGGGGTAAGG ACCATTACCA CAGGCGCCCC
 801 CGTCACATAC TCTACCTATG GCAAGTTCT TGCCGATGGT GGTTGCTCTG
 851 GGGGCGCTTA TGACATCATA ATATGTGATG AGTGCCATTG AACTGACTCG
 901 ACTACAATCT TGGGCATCGG CACAGTCCTG GACCAAGCGG AGACGGCTGG
 951 AGCGCGGCTT GTCGTGCTCG CCACCGCTAC GCCTCCGGGA TCGGTCACCG
 1001 TGCCACACCC AAACATCGAG GAGGTGGCCC TGTCTAATAC TGGAGAGATC
 1051 CCCTTCTATG GCAAAGCCAT CCCCATTGAA GCCATCAGGG GGGGAAGGCA
 1101 TCTCATTTC TGTCAATTCCA AGAAGAAAGTG CGACGAGCTC GCGCAGAAC
 1151 TGTCAAGGCCT CGGAATCAAC GCTGTGGCGT ATTACCGGGG GCTCGATGTG
 1201 TCCGTCATAC CAACTATCGG AGACGTCGTT GTCGTGGCAA CAGACGCTCT
 1251 GATGACGGGC TATACGGCG ACTTTGACTC AGTGATCGAC TGTAACACAT
 1301 GTGTCAACCA GACAGTCGAC TTCAGCTTGG ATCCCCACCTT CACCATTGAG
 1351 ACGACGACCG TGCCTCAAGA CGCAGTGTGCG CGCTCGCAGC GGCAGGGTAG
 1401 GACTGGCAGG GGTAGGAGAG GCATCTACAG GTTTGTGACT CCGGGAGAAC
 1451 GGCCCTCGGG CATGTTCGAT TCCTCGGTCC TGTGTGAGTG CTATGACGCG
 1501 GGCTGTGCTT GGTACGAGCT CACCCCGCC GAGACCTCGG TTAGGTTGCG
 1551 GGCCTACCTG AACACACCAAG GGTTGCCCGT TTGCCAGGAC CACCTGGAGT
 1601 TCTGGGAGAG TGTCTTCACA GGCCTCACCC ACATAGATGC ACACCTCTTG
 1651 TCCCAAGACCA AGCAGGCAGG AGACAACCTTC CCCTACCTGG TAGCATACCA

FIG. 2A

4/92

1701 AGCCACGGTG TGCGCCAGGG CTCAGGCCAC ACCTCCATCA TGGGATCAAA
 1751 TGTGGAAGTG TCTCATACGG CTGAAACCTA CGCTGCACGG GCCAACACCC
 1801 TTGCTGTACA GGCTGGGAGC CGTCCAAAAT GAGGTACCC TCACCCACCC
 1851 CATAACCAA TACATCATGG CATGCATGTC GGCTGACCTG GAGGTCGTCA
 1901 CTAGCACCTG GGTGCTGGTG GGCGGAGTCC TTGCAGCTCT GGCCGCGTAT
 1951 TGCCTGACAA CAGGCAGTGT GGTCAATTGTG GGTAGGATTA TCTTGTCCGG
 2001 GAGGCCGGCT ATTGTTCCCG ACAGGGAGTT TCTCTACCAAG GAGTTCGATG
 2051 AAATGGAAGA GTGCCCTCG CACCTCCCTT ACATCGAGCA GGGAAATGCAG
 2101 CTCGCCGAGC AATTCAAGCA GAAAGCGCTC GGGTTACTGC AAACAGCCAC
 2151 CAAACAAGCG GAGGCTGCTG CTCCCGTGGT GGAGTCCAAG TGGCGAGCCC
 2201 TTGAGACATT CTGGCGAAG CACATGTGGA ATTTCATCAG CGGGATACAG
 2251 TACTTAGCAG GCTTATCCAC TCTGCCTGGG AACCCCGCAA TAGCATCATT
 2301 GATGGCATTACAGCCTCTA TCACCAGCCC GCTCACCACC CAAAGTACCC
 2351 TCCTGTTAA CATCTTGGGG GGGTGGGTGG CTGCCCAACT CGCCCCCCCC
 2401 AGCGCCGCTT CGGCTTCGT GGGCGCCGGC ATCGCCGGTG CGGCTGTTGG
 2451 CAGCATAGGC CTTGGGAAGG TGCTTGTTGA CATTTCTGGCG GGTTATGGAG
 2501 CAGGAGTGGC CGGGCGCTC GTGGCCTTCA AGGTCAATGAG CGGGAGATG
 2551 CCCTCCACCG AGGACCTGGT CAATCTACTT CCTGCCATCC TCTCTCCTGG
 2601 CGCCCTGGTC GTGGGGGTG TGTTGTCAGC AATACTGCGT CGACACGTGG
 2651 GTCCGGGAGA GGGGGCTGTG CAGTGGATGA ACCGGCTGAT AGCGTTCGCC
 2701 TCGCGGGGTA ATCATGTTTC CCCCACGCAC TATGTGCCTG AGAGCGACGC
 2751 CGCAGCGCGT GTTACTCAGA TCCTCTCCAG CCTTACCATC ACTCAGCTGC
 2801 TGAAAAGGCT CCACCAGTGG ATTAATGAAG ACTGCTCCAC ACCGTGTTCC
 2851 GGCTCGTGGC TAAGGGATGT TTGGGACTGG ATATGCACGG TGTTGACTGA
 2901 CTTCAAGACC TGGCTCCAGT CCAAGCTCCT GCCGCAGCTA CCGGGAGTCC
 2951 CTTTTTCTC GTGCCAACGC GGGTACAAGG GAGTCTGGCG GGGAGACGGC
 3001 ATCATGCAA CCACCTGCCAC ATGTGGAGCA CAGATCACCG GACATGTCAA
 3051 AAACGGTCC ATGAGGATCG TCGGGCCTAA GACCTGCAGC AACACGTGGC
 3101 ATGGAACATT CCCCACCAAC GCATACACCA CGGGCCCCCTG CACACCCCTCT
 3151 CCAGCGCAA ACTATTCTAG GGGCTGTGG CGGGTGGCCCG CTGAGGAGTA
 3201 CGTGGAGGTC ACGGGGTGG GGGATTCCA CTACGTGACG GGCATGACCA
 3251 CTGACAACGT AAAGTGCCCA TGCCAGGTT CGGCTCCTGA ATTCTTCACG
 3301 GAGGTGGACG GAGTGCCTG GCACAGGTAC GCTCCGGCGT GCAGGCCTCT
 3351 CCTACGGGAG GAGGTTACAT TCCAGGTGG GCTCAACCAA TACCTGGTTG

FIG. 2B

3401 GGTCACAGCT ACCATGCGAG CCCGAACCGG ATGTAGCAGT GCTCACTTCC
3451 ATGCTCACCG ACCCCCTCCCA CATCACAGCA GAAACGGCTA AGCGTAGGTT
3501 GCCCAGGGGG TCTCCCCCT CCTTGGCCAG CTCTTCAGCT AGCCAGTTGT
3551 CTGCGCCTTC CTTGAAGGCG ACATGCACTA CCCACCATGT CTCTCCGGAC
3601 GCTGACCTCA TCGAGGCCAA CCTCCTGTGG CGGCAGGAGA TGGGGGGAA
3651 CATCACCCGC GTGGAGTCGG AGAACAAAGGT GGTAGTCCTG GACTCTTCG
3701 ACCCGCTTCG AGCGGAGGAG GATGAGAGGG AAGTATCCGT TCCGGCGGAG
3751 ATCCTGCGGA AATCCAAGAA GTTCCCCGCA GCGATGCCCA TCTGGCGCG
3801 CCCGGATTAC AACCTCCAC TGTTAGAGTC CTGGAAGGAC CGGGACTACG
3851 TCCCCTCCGGT GGTGACGGG TGCCCGTTGC CACCTATCAA GGCCCTCCA
3901 ATACCACCTC CACGGAGAAA GAGGACGGTT GTCTAACAG AGTCCTCCGT
3951 GTCTTCTGCC TTAGCGGAGC TCGCTACTAA GACCTTCGGC AGTCCTCGAAT
4001 CATCGGCCGT CGACAGCGGC ACGGCGACCG CCCTTCCTGA CCAGGCTCC
4051 GACGACGGTG ACAAAGGATC CGACGTTGAG TCGTACTCCT CCATGCC
4101 CCTTGAGGGGG GAACCGGGGG ACCCCGATCT CAGTGACGGG TCTTGGTCTA
4151 CCGTGAGCGA GGAAGCTAGT GAGGATGTG TCTGCTGCTC AATGTCCTAC
4201 ACATGGACAG GCGCCTTGAT CACGCCATGC GCTGCGGAGG AAAGCAAGCT
4251 GCCCATCAAC GCGTTGAGCA ACTCTTGCT GCGCCACCAT AACATGGTTT
4301 ATGCCACAAC ATCTCGCAGC GCAGGCCCTGC GGCAGAAGAA GGTCACCTTT
4351 GACAGACTGC AAGTCCTGGA CGACCACTAC CGGGACGTGC TCAAGGAGAT
4401 GAAGGCGAAG GCGTCCACAG TTAAGGCTAA ACTCCTATCC GTAGAGGAAG
4451 CCTGCAAGCT GACGCCCGCA CATTGGCCA AATCCAAGTT TGGCTATGGG
4501 GCAAAGGACG TCCGGAACCT ATCCAGCAAG GCCGTTAACCC ACATCCACTC
4551 CGTGTGGAAG GACTTGCTGG AAGACACTGT GACACCAATT GACACCACCA
4601 TCATGGCAAA AAATGAGGTT TTCTGTGTCC AACCAAGAGAA AGGAGGCCGT
4651 AAGCCAGCCC GCCTTATCGT ATTCCCAGAT CTGGGAGTCC GTGTATGCGA
4701 GAAGATGGCC CTCTATGATG TGGTCTCCAC CCTTCCTCAG GTCGTGATGG
4751 GCTCCTCATA CGGATTCCAG TACTCTCCTG GGCAGCGAGT CGAGTCCCTG
4801 GTGAATACCT GGAAATCAAA GAAAAACCCC ATGGGCTTTT CATATGACAC
4851 TCGCTGTTTC GACTCAACGG TCACCGAGAA CGACATCCGT GTTGAGGAGT
4901 CAATTACCA ATGTTGTGAC TTGGCCCCCG AAGCCAGACA GGCCATAAAA
4951 TCGCTCACAG AGCGGCTTTA TATCGGGGGT CCTCTGACTA ATTCAAAAGG
5001 GCAGAACTGC GGTTATCGCC GGTGCCGCGC GAGCGGCGTG CTGACGACTA
5051 GCTGCGGTAA CACCCTCACA TGTTACTTGA AGGCCTCTGC AGCCTGTCGA

6/92

5101 GCTGCGAAGC TCCAGGACTG CACGATGCTC GTGAACGCCG CCGGCCTTGT
5151 CGTTATCTGT GAAAGCGCGG GAACCCAAGA GGACGCCGGC AGCCTACGAG
5201 TCTTCACGGA GGCTATGACT AGGTACTCTG CCCCCCCCCGG GGACCCGCC
5251 CAACCAGAAT ACGACTTGGA GCTGATAACA TCATGTTCCCT CCAATGTGTC
5301 GGTCGCCAC GATGCATCAG GCAAAAGGGT GTACTACCTC ACCCGTGATC
5351 CCACCACCCC CCTCGCACGG GCTGCGTGGG AAACAGCTAG ACACACTCCA
5401 GTTAACTCCT GGCTAGGCAA CATTATCATG TATGCGCCCA CTTTGTGGC
5451 AAGGATGATT CTGATGACTC ACTTCTTCTC CATCCTTCTA GCACAGGAGC
5501 AACTTGAAA AGCCCTGGAC TGCCAGATCT ACGGGGCCTG TTACTCCATT
5551 GAGCCACTTG ACCTACCTCA GATCATTGAA CGACTCCATG GCCTTAGCGC
5601 ATTTTCACTC CATAGTTACT CTCCAGGTGA GATCAAATAGG GTGGCTTCAT
5651 GCCTCAGGAA ACTTGGGTA CCACCCCTTGC GAGTCTGGAG ACATCGGCC
5701 AGGAGCGTCC GCGCTAGGCT ACTGTCCCAG GGGGGGAGGG CCGCCACTTG
5751 TGGCAAGTAC CTCTTCAACT GGGCAGTGAA GACCAAACTC AAACTCAC
5801 CAATCCCGGC TGCGTCCCAG CTGGACTTGT CCGGCTGGTT CGTTGCTGGT
5851 TACAGCGGGG GAGACATATA TCACAGCCTG TCTCGTGCCC GACCCCGCTG
5901 GTTCATGCTG TGCCTACTCC TACTTTCTGT AGGGGTAGGC ATCTACCTGC
5951 TCCCCAACCG ATAAA

FIG. 2D

7/92

1 GCCACCATGG CCCCATCAC CGCCTACAGC CAGCAGACCC GCGGCCTGCT
 51 GGGCTGCATC ATCACCAAGCC TGACCGGCCG CGACAAGAAC CAGGGTGGAGG
 101 GCGAGGTGCA GGTGGTGAGC ACCGCCACCC AGAGCTTCCT GCCACCTGC
 151 GTGAACGGCG TGTGCTGGAC CGTGTACCAAC GGCGCCGGCA GCAAGACCT
 201 GGCCGGCCCC AAGGGCCCCA TCACCCAGAT GTACACCAAC GTGGACCAGG
 251 ACCTGGTGGG CTGGCAGGCC CCCCGGGCG CCCGCAGCCT GACCCCTGC
 301 ACCTGCGGCA GCAGCGACCT GTACCTGGTG ACCCGCCACG CCGACGTGAT
 351 CCCCGTGCGC CGCCGCGCG ACAGCCCGG CAGCCTGCTG AGCCCCCGCC
 401 CCGTGAGCTA CCTGAAGGGC AGCAGCGCG GCCCCCTGCT GTGCCCCAGC
 451 GCCACGCGCG TGGGCATCTT CCGCGCCGCG GTGTGCACCC GCGGCCTGGC
 501 CAAGGCCGTG GACTTCGTGC CCGTGGAGAG CATGGAGACC ACCATGCGCA
 551 GCCCCGTGTT CACCGACAAC AGCAGCCCCC CCGCCGTGCC CCAGAGCTTC
 601 CAGGTGGCCC ACCTGCACGC CCCCACCGGC AGCGGCAAGA GCACCAAGGT
 651 GCCCCGCCGCC TACGCCGCC AGGGCTACAA GGTGCTGGTG CTGAACCCCA
 701 CGTGGCCGC CACCCCTGGGC TTCGGCCCT ACATGAGCAA GGCCCACGGC
 751 ATCGACCCCA ACATCCGCAC CGGCGTGGCG ACCATCACCA CGGGCGCCCC
 801 CGTGACCTAC AGCACCTACG GCAAGTCCT GGGCGACGGC GGCTGCAGCG
 851 GCGGCGCCTA CGACATCATC ATCTGCGACG AGTGCCACAG CACCGACAGC
 901 ACCACCATCC TGGGCATCGG CACCGTGTG GACCAGGCC AGACCGCCGG
 951 CGCCCCGCTG GTGGTGTGG CCACCGCCAC CCCCCCGGGC AGCGTGACCG
 1001 TGCCCCACCC CAACATCGAG GAGGTGGCCC TGAGCAACAC CGGGAGATC
 1051 CCCTTCTACG GCAAGGCCAT CCCCACATCGAG GCCATCCGCG GCGGCCGCCA
 1101 CCTGATCTTC TGCCACAGCA AGAAGAAGTG CGACGAGCTG GCCGCCAAGC
 1151 TGAGCGGCCT GGGCATCAAC GCGTGGCCT ACTACCGCGG CCTGGACGTG
 1201 AGCGTGATCC CCACCATCGG CGACGTGGTG GTGGTGGCCA CCGACGCCCT
 1251 GATGACCGGC TACACCGGC ACCTCGACAG CGTGATCGAC TGCAACACCT
 1301 GCGTGACCCA GACCGTGGAC TTCAGCCTGG ACCCCACCTT CACCATCGAG
 1351 ACCACCAACG TGCCCCAGGA CGCCGTGAGC CGCAGCCAGC GCGCGGGCG
 1401 CACCGGCCGC GGCGCCGCG GCATCTACCG CTTCGTGACC CCCGGCGAGC
 1451 GCCCCAGCGG CATGTTCGAC AGCAGCGTGC TGTGCGAGTG CTACGACGCC
 1501 GGCTGCGCCT GGTACGAGCT GACCCCGCC GAGACCAGCG TGCGCCTGCG
 1551 CGCCTACCTG AACACCCCCG GCCTGCCGT GTGCCAGGAC CACCTGGAGT
 1601 TCTGGAGAG CGTGTTCACC GGCTGACCC ACATCGACGC CCACTTCCTG
 1651 AGCCAGACCA AGCAGGCCGG CGACAACTTC CCCTACCTGG TGGCCTACCA

FIG. 3A

1701 GGCCACCGTG TCGCCCCGCG CCCAGGCCCC CCCCCCCCAGC TGGGACCAGA
 1751 TGTGGAAGTG CCTGATCCGC CTGAAGCCCCA CCCTGCACGG CCCCACCCCC
 1801 CTGCTGTACC GCCTGGCGC CGTGCAGAAC GAGGTGACCC TGACCCACCC
 1851 CATCACCAAG TACATCATGG CCTGCATGAG CGCCGACCTG GAGGTGGTGA
 1901 CCAGCACCTG GGTGCTGGTG GGC GGCGTGC TGGCCGCCCT GGCCGCCTAC
 1951 TGCCCTGACCA CCGGCAGCGT GGTGATCGTG GGCCGCATCA TCCTGAGCGG
 2001 CCGCCCCGCC ATCGTGCCCC ACCCGGAGTT CCTGTACCAAG GAGTTCGACG
 2051 AGATGGAGGA GTGCCGCCAGC CACCTGCCCT ACATCGAGCA GGGCATGCAG
 2101 CTGGCCGAGC AGTTCAAGCA GAAGGCCCTG GGCCCTGCTGC AGACCGCCAC
 2151 CAAGCAGGCC GAGGCCGCCG CCCCCGTGGT GGAGAGCAAG TGGCGCGCCC
 2201 TGGAGACCTT CTGGGCCAAG CACATGTGGA ACTTCATCAG CGGCATCCAG
 2251 TACCTGGCCG GCCTGAGCAC CCTGCCCGC AACCCCGCCA TCGCCAGCCT
 2301 GATGGCCTTC ACCGCCAGCA TCACCAGCCC CCTGACCACC CAGAGCACCC
 2351 TGCTGTTCAA CATCCTGGGC GGCTGGGTGG CCGCCCAAGCT GGCCCCCCCC
 2401 AGCGCCGCCA GCGCCTCGT GGGCGCCCGC ATCGCCGGCG CCGCCGTGGG
 2451 CAGCATCGGC CTGGGCAAGG TGCTGGTGG CATCCTGGCC GGCTACGGCG
 2501 CCGGCGTGGC CGGCCCGCTG GTGGCCTTCAG AGGTGATGAG CGGCGAGATG
 2551 CCCAGCACCG AGGACCTGGT GAACCTGCTG CCCGCCATCC TGAGCCCCGG
 2601 CGCCCTGGTG GTGGGCGTGG TGTGCGCCGC CATCCTGCAG CGCCACGTGG
 2651 GCCCCGGCGA GGGCGCCGTG CAGTGGATGA ACCGCCTGAT CGCCTTCGCC
 2701 AGCCGCGGCA ACCACGTGAG CCCCCACCCAC TACGTGCCCG AGAGCGACGC
 2751 CGCCGCCCGC GTGACCCAGA TCCTGAGCAG CCTGACCATC ACCCAGCTGC
 2801 TGAAGCGCCT GCACCAAGTGG ATCAACGAGG ACTGCAGCAC CCCCTGCAGC
 2851 GGCAGCTGGC TGCGCGACGT GTGGGACTGG ATCTGCACCG TGCTGACCGA
 2901 CTTCAAGACC TGGCTGCAGA GCAAGCTGCT GCCCCAGCTG CCCGGCGTGC
 2951 CCTTCTTCAG CTGCCAGCGC GGCTACAAGG GCGTGTGGCG CGGCGACGGC
 3001 ATCATGCAGA CCACCTGCCCT GCGGGCGCC CAGATCACCG GCCACGTGAA
 3051 GAACGGCAGC ATGCGCATCG TGGGCCCCAA GACCTGCAGC AACACCTGGC
 3101 ACGGCACCTT CCCCCATCAAC GCCTACACCA CCGGCCCCCTG CACCCCCAGC
 3151 CCGGCCCCCA ACTACAGCCG CGCCCTGTGG CGCGTGGCGC CCGAGGAGTA
 3201 CGTGGAGGTG ACCCGCGTGG GCGACTTCCA CTACGTGACC GGCATGACCA
 3251 CCGACAAACGT GAAAGTGGCCC TGCCAGGTGC CGGCCCGCGA GTTCTTCACC
 3301 GAGGTGGACG CGGTGGCGCT GCACCGCTAC GCGCCCCGCCCT GCGCCCCCT
 3351 GCTGCGCGAG GAGGTGACCT TCCAGGTGGG CCTGAACCAG TACCTGGTGG

9/92

3401 GCAGCCAGCT GCCCTGCGAG CCCGAGCCCG ACGTGGCCGT GCTGACCAGC
 3451 ATGCTGACCG ACCCCCAGCCA CATCACCGCC GAGACCGCCA AGCGCCGCT
 3501 GGCCCCGGC AGCCCCCCCAGC GCCTGGCCAG CAGCAGCGCC AGCCAGCTGA
 3551 GCGCCCCCAG CCTGAAGGCC ACCTGCACCA CCCACCACGT GAGCCCCGAC
 3601 GCCGACCTGA TCGAGGCCAA CCTGCTGTGG CGCCAGGAGA TGGGGGGCAA
 3651 CATCACCCGC GTGGAGAGCG AGAACAAAGGT GGTGGTGCTG GACAGCTTCG
 3701 ACCCCCCTGCG CGCCGAGGAG GACGAGCGCG AGGTGAGCGT GCCCGCCGAG
 3751 ATCCTGCGCA AGAGCAAGAA GTTCCCCGCC GCCATGCCA TCTGGCCCG
 3801 CCCCCACTAG AACCCCCCCC TGCTGGAGAG CTGGAAGGAC CCCGACTACG
 3851 TGCCCCCCGT GGTGCACGGC TGCCCCCTGC CCCCATCAA GGCCCCCCCC
 3901 ATCCCCCCCCC CCCGCCGCAA GCGCACCGTG GTGCTGACCG AGAGCAGCGT
 3951 GAGCAGCGCC CTGGCCGAGC TGGCCACCAA GACCTTCGGC AGCAGCGAGA
 4001 GCAGCGCCGT GGACAGCGGC ACCGCCACCG CCCTGCCCCGA CCAGGCCAGC
 4051 GACGACGGCG ACAAGGGCAG CGACGTGGAG AGCTACAGCA GCATGCCCG
 4101 CCTGGAGGGC GAGCCCGGCG ACCCCGACCT GAGCGACGGC AGCTGGAGCA
 4151 CCGTGAGCGA GGAGGCCAGC GAGGACGTGG TGTGCTGCAG CATGAGCTAC
 4201 ACCTGGACCG GCGCCCTGAT CACCCCCCTGC GCCGCCGAGG AGAGCAAGCT
 4251 GCCCATCAAC GCCCTGAGCA ACAGCCTGCT GCGCCACCA AACATGGTGT
 4301 ACGCCACCAAC CAGCCGCAGC GCCGGCTGC GCCAGAAGAA GGTGACCTTC
 4351 GACCGCCTGC AGGTGCTGGA CGACCAACTAC CGCGACGTGC TGAAGGAGAT
 4401 GAAGGCCAAG GCCAGCACCG TGAAGGCCAA GCTGCTGAGC GTGGAGGAGG
 4451 CCTGCAAGCT GACCCCCCCC CACAGCGCCA AGAGCAAGTT CGGCTACGGC
 4501 GCCAAGGACG TGCGCAACCT GAGCAGCAAG GCCGTGAACC ACATCCACAG
 4551 CGTGTGGAAG GACCTGCTGG AGGACACCGT GACCCCCATC GACACCACCA
 4601 TCATGGCCAA GAACGAGGTG TTCTGCGTGC AGCCCGAGAA GGGCGGCCGC
 4651 AAGCCCGCCC GCCTGATCGT GTTCCCCGAC CTGGGGCGTGC GCGTGTGCGA
 4701 GAAGATGGCC CTGTACGACG TGGTGAGCAC CCTGGCCCCAG GTGGTGTGATGG
 4751 GCAGCAGCTA CGGCTTCCAG TACAGCCCCG GCCAGCGCGT GGAGTTCCTG
 4801 GTGAACACCT GGAAGAGCAA GAAGAACCCC ATGGGCTTCA GCTACGACAC
 4851 CCGCTGCTTC GACAGCACCG TGACCGAGAA CGACATCCGC GTGGAGGAGA
 4901 GCATCTACCA GTGCTGCGAC CTGGCCCCCG AGGCCCCGCCA GGCCATCAAG
 4951 AGCCTGACCG AGCGCCTGTA CATCGGCCGC CCCCTGACCA ACAGCAAGGG
 5001 CCAGAACTGTC GGCTACCGCC GCTGCCGCCGC CAGCGGCGTG CTGACCAACCA
 5051 GCTGCGGCAA CACCCCTGACC TGCTACCTGA AGGCCAGCGC CGCCTGCCGC

FIG. 3C

10/92

5101 GCCGCCAACG TGCAGGACTG CACCATGCTG GTGAACGCCG CCGGCCTGGT
5151 GGTGATCTGC GAGAGGCCG GCACCCAGGA GGACGCCGCC AGCCTGCCGCG
5201 TGTTCACCGA GGCCATGACC CGCTACAGCG CCCCCCCCCGG CGACCCCCCC
5251 CAGCCCCAGT ACGACCTGGA GCTGATCACC AGCTGCAGCA GCAACGTGAG
5301 CGTGGCCCAC GACGCCAGCG GCAAGCGCGT GTACTACCTG ACCCCGCCACC
5351 CCACCAACCCC CCTGGCCCGC GCCGCCTGGG AGACCGCCCG CCACACCCCC
5401 GTGAACAGCT GGCTGGCAA CATCATCATG TACGCCCCA CCCTGTGGC
5451 CCGCATGATC CTGATGACCC ACTTCTTCAG CATCCTGCTG GCCCAGGAGC
5501 AGCTGGAGAA GGCCCTGGAC TGCCAGATCT ACGGCGCCTG CTACAGCATC
5551 GAGCCCCCTGG ACCTGCCCA GATCATCGAG CGCCTGCACG GCCTGAGCGC
5601 CTTCAGCCTG CACAGCTACA GCCCCGGCGA GATCAACCGC GTGGCCAGCT
5651 GCCCTGCACAA GCTGGCGTG CCCCCCTGC GCGTGTGGCG CCACCGCGCC
5701 CGCAGCGTGC GCGCCCGCCT GCTGAGCCAG GCGGGCCCGCG CCGCCACCTG
5751 CGGCAAGTAC CTGTTCAACT GGGCCGTGAA GACCAAGCTG AAGCTGACCC
5801 CCATCCCCGC CGCCAGCCAG CTGGACCTGA GCGGCTGGTT CGTGGCCGGC
5851 TACAGCGCG GCGACATCTA CCACAGCCTG AGCCGCCGCC GCCCCCGCTG
5901 GTTCATGCTG TGCCCTGCTGC TGCTGAGCGT GGGCGTGGGC ATCTACCTGC
5951 TGCCCCAACCG CTAAA

FIG. 3D

11/92

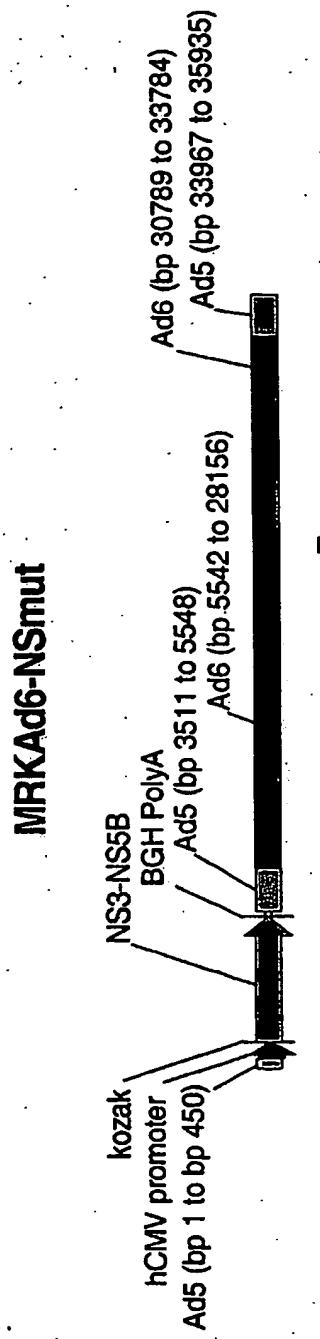


FIG. 4A

12/92

FIG. 4B

13/92

3361 gcccggcaat tcaaggcaga aagcgctcgaa ttactgc aaaa cagccacccaa acaaggcgag
 3421 gctgctgctc ccgtgggtgaa gtccaaagtgg cgagccctt agacattctg ggcgaaggcac
 3481 atgttggatt tcatcagcg gatacagtac tttagcaggt tatccactt ctgcggaa
 3541 ccccaatag catcattgtat ggcattcaca gcctctatca ccagcccgct caccacccaa
 3601 agtacccctcc tggtaacat ctgggggggg tgggtggctg cccaaatcgc cccccccagc
 3661 gccgttcgg ctttcgtgg cgccggcata gcccgtcggt ctgttggcag cataggcctt
 3721 gggaaagggtgc ttgtggacat tctgggggt tatggagcag gagtggccgg cgccgtcg
 3781 gccttcaagg tcatgagcg cgagatgccc tccaccgagg acctgtcaa tctacttc
 3841 gccatcctct ctccctggcgc cctggcgtc ggggtcggt gtgcagcaat actgcgtcg
 3901 cacgtgggtc cgggagagg ggctgtgcag tggatgaacc ggctgtatagc gtgcgttc
 3961 cgggtaatc atgttcccc caccactat gtgcctgaga ggcacccgcg agcgcgttt
 4021 actcagatcc tctccagcct taccatcaact cagctgtaa aaaggctcca ccagtggatt
 4081 aatgaagact gctccacacc gtgttccggc tcgtggctaa gggatgtttg gacttgata
 4141 tgcacgggtgt tgactgaccc caagacccgg ctccagtcata agctcctgccc gcaagctaccc
 4201 ggagttccctt ttttcgtg ccaacccggg tacaaggag tctggggggg agacggcata
 4261 atgcacaaacca cctgcccattt tggagcacag atcaccggac atgtcaaaaaa cgggtccatg
 4321 aggatcgctg ggcctaagac ctgcaccaac acgtggcatg gaacatccc catcaacgc
 4381 tacaccacgg gcccctgcac acccttcata ggcggcaactt attctaggc gctgtgggg
 4441 gtggccgtg aggatgtacccgg gggatgtacgg cgggtggggg attttccacta cgtgacgggg
 4501 atgaccactt acaacgtaaa tgcccatgc cagggtccgg ctccatgattt cttcacggag
 4561 gtggacggag tgccgttgca caggtacgct cggccgtgca ggcctctt acgggaggag
 4621 gttacattcc aggtcggttca accacaatac ctgggtgggt cacagctacc atgcgagccc
 4681 gaaccggatg tagcagtgtc cacttccatg ctcaccgacc cctccacat cacaggcaga
 4741 acggctaagc gttagttggc caggggtctt ccccccctt tggccagtc ttcagctag
 4801 cagttgtctg cgccttcctt gaaggcgaca tgcactacc accatgtctc tccggacgct
 4861 gacccatcg aggccaaacctt cctgtggcgg caggagatgg ggcggaaacat caccggcg
 4921 gagtcggaga acaagggtgtt agtccctggac tctttcgacc cgttcgagc ggaggaggat
 4981 gagagggaaat tatccgtttcc ggcggagatc ctgcggaaat ccaagaatgg ccccgccagc
 5041 atgcacccatggggcgccc ggattacaac cctccactgt tagatcttgc gaaaggacccg
 5101 gactacgtcc ctccgggtgt gcacgggtgc cctggccac ctataaaggc ccttccatata
 5161 ccacccatccac ggagaaagag gacgggtgtc taaacagatg cctccgtgtc ttctgcctt
 5221 gccggagctcg ctactaagac cttccggcgc tccgaatcat cggccgtcgca cagccggc
 5281 ggcacccggcc ttccatgacca ggcctccgac gacgggtgaca aaggatccga ctgttgc
 5341 tactccatca tgccccccctt tgagggggaa cccggggacc cccatctcag tgacgggtct
 5401 tggcttaccg tgagcgagga agctagttagt gatgtcgct gctgtcaat tccctacaca
 5461 tggacaggcg ctttgcatac gccatgcgtc gccggaggaaa gcaagctgccc catcaaccc
 5521 ttgagcaact ctttgcgtcg ccaccataac atggtttatg ccacaaacatc tccgcaccc
 5581 ggcctcgccgca agaagaaggatc caccatgcgca agactgcgaa tccctggacca ccaactacc
 5641 gacgtgtca aggagatgaa ggcgaaggcg tccacagtttta aggtctaaactt cctatccgt
 5701 gaggaagccct gcaagctgac gcccacat tcggccaaat ccaatgtttgg ctatggggca
 5761 aaggacgtcc ggaacccatc cagaaggcc gttaaaccaca tccactccgt gtggaaaggac
 5821 ttgcgtggaaatcgtgtac accaattgtacccatca tgccaaaaaa tgagggtttc
 5881 tgcgtccaaatc cagagaaagg aggccgtaaatc ccagccggcc ttatctgttccatctg
 5941 ggagttccgttgc tgcgtggaaatc gatggccctt tgcgtggatgg tccctcgggt
 6001 gtgtatggctt cctccatcgttgc attccatgtac tccctggcc gacggatgtca gttccctgg
 6061 aataccctggaa aatcaaaggaa aacccatgtt ggcctttcat atgacactcg ctgttgc
 6121 tcaacccgtca ccgagaaccaatc catccgtgtt gaggagtttcaat tttaccaatg ttgtactt
 6181 gccccccggaa ccagacaggc cataaaatcg ctcacagagc ggcttataat cgggggtct
 6241 ctgactaattt caaaaggccgaa gaaatgcgtt tgcgtggatgg gccggccgag cggccgtcg
 6301 acgactatgtt gccgttacac cctccatgttacttgcagg cctctgcac gtcgtcgat
 6361 gcaagctcc aggactgcac gatgtcggt aacggccggc gcttgcgt tatctgtgaa
 6421 agcgccggaa cccaaaggatc ggcggccgac ctacgactt tcaacggaggc tatgtactt
 6481 tacttgccttccccc cccggggaaatc cccggccaaatc cagaatacg acttggagatc gataacatca
 6541 tggccatccatgttgcggat gcatcgatc gatcgacccaaatc cggccatccatc
 6601 cgtgatccca ccacccatgtt ggcacggc gcttgcgtt cagctagacca cactccatc

FIG. 4C

14/92

FIG. 4D

15/92

9961 ttcctgaagg ggggctataa aagggggtgg gggcgcggtc gtcctcaactc tcttccgcatt
 10021 cgctgtctgc gaggggccacg tggtgggtg agtactccct ctcaaaaagcg ggcattgactt
 10081 ctgcgctaag attgtcagg tccaaaaacg aggaggattt gatattcacc tggcccgccg
 10141 ttagtgcctt gagggtggcc gcgtccatct ggtcagaaaa gacaatctt ttgttgtcaa
 10201 gcttgggtggc aaacgaccgg tagagggcgt tggacagcaa cttggcgatg gagcgcagg
 10261 tttgggtttt gtcgcgatcg ggcgcctt gggccgcatt gtttagctgc acgtattcgc
 10321 ggcgaacgcgca cggccattcg gggaaagacgg tgggtgcgtc gtcgggactt aggtgcacgc
 10381 gccaaaccgcg gttgtgcagg gtgacaagggt caacgcgtt ggttgcgtt cccgcgttaggc
 10441 gctcggtggc ccagcagagg cggccgcctt tgcgcgagca gaatggcggt agtgggtcta
 10501 gctgcgtctc gtccgggggg tctgcgttca cggtaaagac cccgggcagec aggccgcgt
 10561 cgaagtagtc tatcttgcatt ctttgcattt ctaggcctt ctgcattgcg cggggggccaa
 10621 ggcgcgcgtc gtatgggtt agtggggac cccatggcatt ggggtgggtt aagcggggagg
 10681 cgttacatgcc gcaaaatgtcg taaacgtaga ggggtcttctt gatgttccaa agatatgttag
 10741 ggttagcatct tccaccgcgg atgtggcgc gacgttatact gttatgttgc tgcgaggagg
 10801 cgaggagggtc gggaccggg ttgttacggg cgggtgttca tgcgttccaa actatcttcc
 10861 tgaagatggc atgttgcattt gatgtatgg ttggacgcgtc gaagacgtt aagctggcgt
 10921 ctgtgagacc taccgcgttca cgcacggagg agggttgcgtt gtcgcgcgtc ttgttgcacca
 10981 gctccgggtt gacctgcacg tctaggcgcc agttagtccat ggttgcgtt atgtatgtcat
 11041 actttatcttgc tccctttttt ttccacatgtt cgcgggtttagt gacaaactct tcgcggctt
 11101 tccagttactc ttggatcggtt aaccgcgttgc cttccgttac gtaagagcctt agcatgtaga
 11161 actgggttgc ggccttggtag ggcgcacgc ctttttcttcc ggttgcgtt gatgttgcgt
 11221 cggcccttccg gaggcgagggtt tgggttgcgtt ctttgcgtt gacaaactct tcgcggctt
 11281 actgggttattt gaaatgttgc tgcgttgcgtt cgcgggtttagt gacaaactct tcgcggctt
 11341 gtttttggc acgcgggtt ggcaggcggtt aggttgcattt gtttgcgtt gatgttcccg
 11401 cgcgaggcat aaatgttgcgtt gtttgcgtt ggggttcccg cacccgttccaa cgggttgcgtt
 11461 ttacctggc ggcgaggcat gtcgttgcgtt ggggttcccg cacccgttccaa cgggttgcgtt
 11521 gtttcaagaa ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 11581 gcttcgttccg gtagcttgcgtt cccttgcgtt aagggttcccg gtttgcgtt gatgttcccg
 11641 aagcgacgaa ggcgggtt ggcagggtt cccttgcgtt aagggttcccg gtttgcgtt gatgttcccg
 11701 tcctaaactg ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 11761 cttgttccca ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 11821 gctcatctcc ggcgaacttcc atgaccagca tgaagggttcc gaggcttcc gtttgcgtt gatgttcccg
 11881 ccatccaagt ataggcttcc acatgttcc gtttgcgtt gacaaagag acgcttgcgtt ggggttcc
 11941 agccgatcg gtttgcgtt gacaaagag acgcttgcgtt ggggttcc gtttgcgtt gatgttcc
 12001 gaaagtagaa gtttgcgtt gacaaagag acgcttgcgtt ggggttcc gtttgcgtt gatgttcc
 12061 agtactggca ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12121 caaggaagca ggttgcgtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12181 ctttgcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12241 ccacgcgcgtt ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12301 catcgccgtt ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12361 gcttcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12421 tgatgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12481 ggcggactac ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12541 ctaaaaagcg ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12601 agggggcagg ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12661 gcttgcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12721 gacggggcccg ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12781 gacggggcccg ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12841 ggcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12901 ggcggccagg ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 12961 gtttgcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 13021 cgcggactac ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 13081 gtagtttgcgtt ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 13141 cgttgcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt
 13201 ggcgttcc ggcgggtt ggcagggtt cccttgcgtt aaggcaattt ttttttttccaa tgcgttgcgtt

FIG. 4E

16/92

13261 gatgagctcg gcgacagtgt cgccacaccc gcgctcaaag gctacagggg cctcttcttc
 13321 ttcttcatac tccttctcca taaggcctc cccttcttct tcttctggcg gccgtggggg
 13381 aggggggaca cggccgcac gacggcgcac cgggaggcgg tcgacaaaagc gctcgatcat
 13441 ctccccgcgg cgacggcgcac tggctcggt gacggcgcgg ccgttctcgc gggggcgcag
 13501 ttgaaagacg ccgcccgtca tggccgggtt atgggttggc ggggggctgc cgtcgccgcag
 13561 ggatacggcg ctaacgatgc atctcaacaa ttgttgcgtta ggtactccgc caccgaggga
 13621 cctgagcgag tccgcatcga cccgatcgaa aacctctcg agaaaggcgt ctaaccagtc
 13681 acatcgcaaa ggtaggctga gcaccgtggc gggccgcag gggccgggt cggggttgtt
 13741 tctggccggag gtgtcgctga tgatgttaat aaagtaggcg gtcttgagac ggcggatggt
 13801 cgacagaagc accatgtcct tgggtccggc ctgctgaatg cgcaggcgt cgccatgcc
 13861 ccaggctcg ttttgcacatc ggcgcaggc tttgttagtag tcttgcacatc gccttctac
 13921 cggcacttct tcttctccct cctttgtcc tgcacatctt gcatctatcg ctgcggcggc
 13981 ggcggagttt ggccgttagt ggccgcctct tcctcccatg cgtgtgaccc cgaagccct
 14041 catcgctga agcaggggcca gtcggcgcac aacgcgcctcg gctaataatgg cctgctgcac
 14101 ctgcgtgagg gtagactgga agtcgtccat gtccacaaaag cgggtgtatg cggccgtgtt
 14161 gatgggttaa gtgcagttgg ccataacgga ccagtttaacg gtctgggtac cgggctgcga
 14221 gagctcggtg tacctgagac gcgagtaaagc ctttgcgtca aagacgtagt ctttgcgt
 14281 ccgcaccagg tactggtatacc ccacaaaaaa gtgcggcggc ggctggcgtt agagggccca
 14341 gctgttaggtg gccggggcgc cggggccggag gtcttccaaac ataaggcgtt gatatccgt
 14401 gatgtacccgt gacatccagg tgatggccgc ggcgggtggg gaggccgcgc gaaagtca
 14461 gacgcgggtc cagatgtgc gcagcggcaaa aagtgcgtt atggcgggaa cgctctggc
 14521 ggtcaggcgc ggcgcgttgc tgacgtctca gaccgtgc aaaggagagcc tgtaagcggg
 14581 cacttcccg tggctcggtt gataaattcg caagggtatc atggcggacg accggggttc
 14641 gaaccccgga tccggccgtc cgcgtgtatc catgcgttta cgcggcgtt gtcgaaccca
 14701 ggtgtgcac gtcagacaac gggggagcgc tcctttggc ttccctccag gcgcggcgg
 14761 tgctcgctca gcttttttgg ccactggccg cgcgcggcgt aagcgtttag gctggaaagc
 14821 gaaagcatra agtggctcgc tccctgttagc cggagggtta ttttccaagg gttgagtcgc
 14881 gggacccccc gttcgagttc cggggccggc ggactgcggc gaacgggggtt ttgcctcc
 14941 gtcatgcacaa accccgtttt caaattccctc cggaaacagg gacgagcccc tttttgtct
 15001 ttcccagatg catccgggtc tgcggcagat ggcggccctt cctcagcgc gcaagagac
 15061 agagcagcg gacatgcgaa gggcaccctc cccttctctt accgcgtcgt gaggggcaac
 15121 atcccggtct gacgcggcgg cagatgtgtaa ttacgaaccc cgcggcgcgc ggacccggca
 15181 ctacttggac ttggaggagg gcgaggggctt ggcgcggcgtt ggagcggccctt ctcctgagcg
 15241 acacccaaagg gtcagctgtc agcgtgacac ggcgcggcgtt tacgtggcgc ggcagaaccc
 15301 gtttgcgac cgcggggggaggaggccgaa ggagatgcgg gatcgaaagt tccatgcagg
 15361 ggcgcgttgc cggcatggcc tgaaccgcga gcggttgcgt cgcggaggagg actttgagcc
 15421 cgacgcgcgg accgggattttt gtcggcgcgc ggcacacgtt ggcggccggc acctggtaac
 15481 cgcgtacgag cagacgggtt accaggagat taactttcaa aaaagcttta acaaccacgt
 15541 ggcgcacgtt gtcggcgcgc aggagggtggc tataggactg atgcacatgtt gggactttgt
 15601 aagcgcgtt gacaaaacc caaatagcaa gccgctcatg ggcgcgttgc tccttatagt
 15661 gcagcacagc agggacaacg aggcatcgtt ggtgcgttgc taaacatag tagagccgaa
 15721 gggccgttgc ctgcgttgcatt tgataaacat tctgcagatc atagggtgtc aggagcgc
 15781 cttgagccctg gctgacaaagg tggccgcatt taacttatttcc atgcctcgtc tggcaagtt
 15841 ttacgcccgc aagatataacc atacccttta cgttccatc gacaaggagg taaagatcg
 15901 ggggttctac atgcgtatgg cgcgttgcgtt gtttacccgtt agcgcacgacc tggcgttta
 15961 tcgcaacgag cgcgttgcatac aggccgtgt cgttgcgttgc ggcgcggcgc tcagcgacc
 16021 cgagctgttgc cacaaggcttca aaggccctt ggcgttgcgttgc ggcgcggcgc atagagaggc
 16081 cgagtcctac ttgcgttgcgtt ggcgttgcgttgc ggcgttgcgttgc ccaagccgc ggcggcc
 16141 ggcagctggg gccggacgtt ggcgttgcgttgc ggcgttgcgttgc cgcgttgcgttgc acgtcgcc
 16201 cgtggaggaa tatgtacggagg acgttgcgttgc ggcgttgcgttgc ggcgttgcgttgc
 16261 gatgttcttgc atcagatgttgc gcaagacgcac acggaccggc cgggtgcggc ggcgttgc
 16321 agccagccgtt cccggccctaa ctccacggac gactggcgc gggccgttgc ggcgttgc
 16381 tcgctgacttgc cgcgcaccc ttgcgttgcgttgc cggcaggcgc cgcaggccaa cccggcttcc
 16441 gcaattctgg aaggcgttgcgttgc gcaaaacccca cgcacgcacaa ggtgtgttgc
 16501 atcgtaaaccgcgaaacaggccgatccggcccgatggcccgatccggcccgatccggcccgatccggcc

FIG. 4F

17/92

16561 gacgcgctgc ttcagcgctg ggctcgttac aacagcagca acgtgcagac caacctggac
 16621 cggctgggg gggatgtgcg cgaggccgtg ggcgcgcgtg agcgcgcgc gacgcaggcc
 16681 aacccggct ccatgggtgc actaaacgc ttcctgaga cacagccgc caacgtgcgc
 16741 cggggacagg aggactacac caacttgcg agcgcactgc ggctaattgt gactgagaca
 16801 ccccaaaatg aggtgtatca gtcggggca gactatccc tccagaccag tagacaaggc
 16861 ctgcagaccg taaaacctgag ccaggcttc aagaacttgc aggggctgtg ggggggtgcgg
 16921 gtcggccacag ggcgcgcgc gaccgtgtc agcttgcgta cgcccaactc ggcgcgtgt
 16981 ctgcgtctaa taggcgcctt cacggacagt ggcgcgtgt cccgggacac ataccttaggt
 17041 cactgtcga cactgtaccc cgaggccata ggtcaggcgc atgtggacga gcatacttc
 17101 caggagatta caagtgttag cgcgcgcgtg gggcaggagg acacgggcag cctggagcca
 17161 accctgaact acctgtctgac caaccggccg caaaaaaatcc ctcgttgca cagttaaac
 17221 agcgaggagg agcgcattt ggcgcgtgt cagcagacg tgagccttaa ctcgtatgcgc
 17281 gacggggtaa cggcccgacgt ggcgcgtggac atgaccgcgc gcaacatgga accgggcgt
 17341 tatgcctcaa accggccgtt tatcaatgcgc ctaatggact acttgcacgc cggggccggc
 17401 gtgaaccccg agtatttccac caatgcctc ttgaacccgc actggctacc gccccctgt
 17461 ttctacaccc ggggattcga ggtgcccgg ggttaacgtg gattcttcg ggacgacata
 17521 gacgacgcg tttttccccc gcaaccgcg accctgttag agttgcaaca acgcgacgc
 17581 gcagaggccg cgcgtcgaaa gggaaatcc cgcaggccaa gcaacatggc cgtatctggc
 17641 gtcggccccc cgcgtcgaaa tgcttagtgc ccatttccaa gcttgatagg gtctttaacc
 17701 agcactcgca ccaccggccc ggcgcgtgt ggcgcaggagg agtacccaaa caactcgctg
 17761 ctgcagccgc agcgcgaaaa gaaacctgcct cccgcgttcc ccaacaacgg gatagagagc
 17821 ctatgtgaca agatgagtag atggaaagacg tatgcgcagg agcacaggga tggcccccgc
 17881 ccgcgcgcgc ccaccgcgtc tcaaaaggcac gaccgtcgc ggggtctgggt tggggaggac
 17941 gatgactcgg cagacgacag cagcgtttt gattttggag ggagtggcaaa cccgtttgca
 18001 cacccgcgc ccaggctggg gagaatgtt taaaaaaaag catgtgcaaa aaaaaaaaaac
 18061 tcaccaaggc catggcaccgc agcgttgggtt ttcttgcatt ccccttagta tgcggcgcgc
 18121 ggcgtatgtat gaggaaaggc ctcctccctc taccggagac gttggagcgg cggcgcgcagt
 18181 ggcgcgcgc ctgggttccac ctttcgtgc tccctggac cccgcgttgc tgcctccgcgc
 18241 gtacccgtgg cttaccgggg ggagaaacag cttccgttac tctgtgttgc cacccttatt
 18301 cgacaccacc cgtgtgtacc ttgtgaccaaa caagtcaacg gatgtggcat ccctgaacta
 18361 ccagaacgcac cacagcaact ttctaaaccac ggtcattcaa aacaatgact acagcccg
 18421 ggaggcaagc acacagacca tcaatcttgc cgaccggcgtc cactggggcg ggcacactgaa
 18481 aaccatctgc cataccaaca tgccaaatgt gaaacgagttc atgtttacca ataagttaa
 18541 ggcgcgggtg atgggtgcgc gtcgcgttac taaggacaaa caggtggagc tggaaatacga
 18601 gtgggtggag ttcacgcgtc cccggggcaaa ctactcccgag accatgacca tagaccttat
 18661 gaacaacgcgc atcgtggagc actactgaa agtgggcagg cagaacgggg ttctggaaag
 18721 cgacatcgcc gtaaaatgtt acaccgcgaa cttcagactg gggtttggacc cagtcaactgg
 18781 tcttgcatt cctggggat atacaaaacgcg agccttccat ccagacatca ttttgcgtgcc
 18841 aggatgcggg gtggacttca cccacagccg cctgagcaac ttgtgggca tccgcacaggc
 18901 gcaacccttc caggagggtt ttagatccat ctcacgtac ctggagggtg gtaacattcc
 18961 cgcactgttg gatgtggacg cttaccaggc aagcttggaa gatgcacccg aacaggccgg
 19021 gggggccgc ggcggccggca acaacgtgg cgcggccgcg gaagagaact ccaacgcgc
 19081 agtcgcggca atgcagccgg tggaggacat gaaacgatcat gccattcgcc ggcacaccc
 19141 tgccacacgg gccggaggaga agcgcgtgc ggcgcggccg gggccgaaag ctggccccc
 19201 cgcgtccggag gctgcacaac ccgagggtgc gaaaggctcg aagaaacccgg tgattaaacc
 19261 cctgacagag gacagcaaga aacgcgttca caacctaata agcaatgaca gcacccatc
 19321 ccagttccgc agtcgttacc ttgcataccaa ctacggccgc cctcaggccg gatccgc
 19381 atggaccctg ctttgcactc ctgacgtac ctgcggctcg gaggcgttactggcgtt
 19441 gcccgcacatg atgcaagacc ccgtgacccctt ccgcgtccacg cgcgcacatca gcaacttcc
 19501 ggtgggtggc gccgagctgt tgccctgcgca ctccaaagagc ttctacaacg accaggccgt
 19561 ctactccctc ctcatccgcg agtttacccctc tctgaccac gtttcaatc gctttccca
 19621 gaaccagatt ttggcgccgc cccggccccc caccatccacc accgtcgtg aaaaacgttcc
 19681 tgctctcaca gatcacggga cgcgtaccgt ggcgcacagc atcgaggagg tccagcgt
 19741 gaccattact gacgcgcac gccgcacccgt cccctacgtt tacaaggccc tgggcataatg
 19801 ctcggccgcgc gtcctatcga gcccactt ttgagcaagc atgtccatcc ttatatcgcc

FIG. 4G

18/92

19861 cagcaataac acaggctggg gcctgcgtt cccaagcaag atgtttggcg gggccaagaa
 19921 gcgctccgac caacacccag tgcgcgtgcg cgggcactac cgcgcgcctt gggcgcgca
 19981 caaacgcggc cgcactggc gcacccacgt cgatgacgc atcgacgcgg tgggtggagga
 20041 ggcgcgcaac tacacgcccc cgcggccgc agtgcgcacc gtggacgcgg ccattcagac
 20101 cgtggtgcgc ggagcccgcc gctacgctaa aatgaagaga cggcggaggg cgtagcactg
 20161 tcgcacccgc cggcggcccg gcaactgcgc ccaacgcgcg gccggggccc tgcttaaccg
 20221 cgcacgtgcg accggccgac gggcgccat gcgagccgt cgaaggctgg cgcgggtat
 20281 tgtcaactgtg ccccccaggc ccaggccgacg agcggccgcg gcagcagccg cggccattag
 20341 tgctatgact cagggtgcga gggcaacgt gtactgggtg cgcgactcggtt tagcggccct
 20401 ggcgcgtgccc gtgcgcaccc gccccccgcg caactagatt gcaataaaaaa actacttaga
 20461 ctcgtactgt tgtatgtatc cagcggccgc ggcgcgcaccc gaaagctatgtt ccaagcgca
 20521 aatcaaagaa gagatgcctt aggtcatgcg gccggagatc tatggccccc cggaaagagga
 20581 agaggcaggat tacaagcccc gaaagctaaa gccgggtcaaa aaaaaaaaga aagatgatga
 20641 tgatgatgaa cttgacgcag aggtgaaact ttgcacgcg accgcggccca ggcacgggt
 20701 acagtggaaa ggtcgcacgc taagacgtt ttgcacgcg ggcaccccg tagtctttac
 20761 gcccccgtgag cgcctccaccc gcacccatc ggcgcgtgtat gatgagggtt acggcgacga
 20821 ggacactgtt ggcacccggc acggccgcct cggggagttt gcctacggaa agcggcataa
 20881 ggacatgtct ggcgttgcgc tggacgaggg caacccaaaca cctagctaa agcccggtgac
 20941 actgcagcag gtgcgtcccg cgctgcacc gcctgaagaa aagcggccca ggcacgggt
 21001 gtcgtggtag ttggcacccca cgcgtcagct gatggtaccc aagcgtcagc gactggaaaga
 21061 tgtcttggaa aaaatgaccg tggagcctgg gctggagccc gagggtccgcg tgcggccaat
 21121 caagcagggt gCACCGGAC tgggtgcgc gaccgtggac gttcagatac ccaccaccc
 21181 tagcactagt attgcactg ccacagaggg catggagaca caaaatgttcc cgggttgcctc
 21241 ggcgggtggca gatgcggccg tgcaggccgc cgctgcggcc ggcgttcaaga cctctacgga
 21301 ggtgcaaaacg gacccgtgga tgcgtgttgcg cgcgttccccc cgggttccgc ggcgttcaag
 21361 gaagtacggc gccgcgcagcg cgctactgcg cgaatatgccc ctacatcctt ccacgcgc
 21421 tacccccggc tgcgtgttgcg acacatccg ccccaagaaga cgagcaacta cccgacgc
 21481 aaccaccaact ggaacccggcc gccgcgtcg cgcgtcccg cccgtgttgcg ccccgatttc
 21541 cgtgcgcagg gtggctcgcg aaggaggcag gacccgttgcg ctgcacacag cgcgttacca
 21601 ccccgacatc gttttaaaacg cgggttttgcg gtttgcgc gatatggccc tcacctgcgc
 21661 cctccgtttc cgggtccgg gattccgagg aagaatgcac cgtaggaggg gcatggccgg
 21721 ccacggccctg acggggccgc tgcgtgtgc gcaccacccg cggcggccgc cgtcgcacc
 21781 tcgcacgcgc ggcgtatcc tgcgttgcgc tattccactg atgcggccgg cgttggcgc
 21841 cgtgcgcggc attgcacccg tggcgttgcg ggcgcagaga cactgattaa aacaaggta
 21901 catgtggaaa aatcaaaata aaagtctggc ctctcacgc cgctttgttgcg ttaactatt
 21961 ttgttagaattt gaaagatca accttgcgtc actggccccc cgacacggct cgcggccgtt
 22021 catggggaaac tggcaagata tcggcaccag caatgatgac ggtggccct tcaactgggg
 22081 ctcgtgtgg agcggcatc aaaaatccgg tttccggcgtt aagaactatg gcaaaaagg
 22141 ctggaaacgc agcacagcc agatgttgcg ggacaagttt aaagagcaaa atttccaaaca
 22201 aaagggtggta gatggcttgcg cctctggcat tagccccgggtt gttggacccgc ccaaccagg
 22261 agtgcaccaat aagattaaaca gtaacgttgcg tccccccctt cccgttagagg agcctccacc
 22321 ggcgtggag acagtgttgcg cagagggccg tggcggaaaag cgtccgcgc acgcacagg
 22381 agaaactctg gtgacgcggaa tagacgagcc tccctcgatc gaggaggcac taaagcaagg
 22441 cctggccacc acccgccca tgcgcgcctt ggctaccggc gtgcgtggcc agcacacacc
 22501 cgtacgcgtc gacccgttgcgc ccccccggccg caccggccgaa acacccgttgc tgcgttgc
 22561 gtccggccgtt gttgtatcc gtcctagccg cgcgtccctt cgcgcgcgc cccgggttcc
 22621 ggcgttgcgtt cggccggccg ccacgtggccg acactgatc gatcgatgggg
 22681 tttgggggtt caatccctga agcggccgcg atgcgttgcg tagctaacgtt gtcgtatgtt
 22741 tgtcatgtat ggcgtccatgt cggccgcaga ggagctgttgc agccgcgcg cggccgcgtt
 22801 ccaagatggc tacccttcg atgtggccgc agtggcttgc catgcacatc tccggccagg
 22861 acggcttgcg gtacccgttgcg cccggccgtt tgcgttgcgc cccggccacc gagacgtact
 22921 tcaggcttgcg taacaaggat agaaacccca cgggtggccgc tacgcacgc gtcggccacc
 22981 accggcttgcg ggcgttgcgttca tccctggatcc cccggccaggat actgcgtact
 23041 cgtaccaaggc ggcgttgcgttca ctagctgtgg gtgataaccg tgcgttgcgc atggcttcca
 23101 cgtactttgcg catccgcggc gtgtggaca gggccctac ttttaagccca tactctggca

FIG. 4H

19/92

23161 ctgcctacaa cgcactggcc cccaagggtg cccccaactc gtgcgagtgg aacaaaatg
23221 aaactgcaca agtggatgct caagaacttg acgaagagga gaatgaagcc aatgaagctc
23281 aggcgcgaga acaggaacaa gctaagaaaa cccatgtata tgcccaggct ccactgtccg
23341 gaataaaaat aactaaagaa ggtctacaaa taggaactgc cgacgcccaca gtagcagggt
23401 cggcaaaaga aatttcgca gacaaaactt ttcaacctga accacaagtg ggagaatctc
23461 aatggaaacg agcggatgcc acagcagctg gtggagggt tctaaaaag acaactccca
23521 tggaaaccctg ctatggctca taagctagac ccaccaattc caacggcggg cagggcgtta
23581 tgggtgaaca aaatgtaaa ttggaaagtc aagtgcgaaat gcaattttt tccacatcca
23641 caaatggccac aaatgaagtt aacaataatac aaccaacatgt ttttttttgc aaccaatcc
23701 taaacatgga aactccagat actcatctt ttatccatggg taaaatgggg gataaaaaatg
23761 ccaaaggatcat gctggacaa caagcaatgc caaacagacc aaattacatt gcttttagag
23821 acaattttat tggctctatg tattacaaca gcacaggtaa catgggtgtc cttgctggtc
23881 aggcatcgca gttgaacgct gtttagatt tgcaagacag aaacacagag ctgtctacc
23941 agcttttgc tgattcaatt ggccgacagaa caagatactt ttcaatgtgg aatcaagctg
24001 ttgacagcta tgatccagat gtcagaattt ttgagaaccg tggactgag gatgagttgc
24061 caaattatttgc ctttcctctt ggtggattt ggattactga cactttcaa gctgtaaaa
24121 caactgtgc taacggggac caaggcaata ctacctggca aaaagattca acattgcag
24181 aacgcaatga aataggggtg gaaaataact ttgcatggg aattaaacctg aatgcacacc
24241 tatggagaaa ttcccttac tccaaatatttgc cgctgtaccc gccagacaag ctaaaataca
24301 accccccacca tggggaaata tctgacaacc ccaacaccta cgactacatg aacaagcgag
24361 tgggtggctcc tgggcttgc gactgctaca ttaaccttgg ggcgcgtgg tctctggact
24421 acatggacaa ctgttatecc ttaaccacc accgcattgc gggcctgcgt taccgttccca
24481 tgggtttggg aaacggccgc tacgtggccct ttcacattca ggtggcccaa aagtttttg
24541 ccattaaaaat cttccctctt ctggccaggct catacacaata tgaatggaaat ttcaggaagg
24601 atgttaacat gttctgcag agctctctgg gaaacgacct tagagttgac gggcttagca
24661 ttaagtttgc cagcatttgc ctttacgcca ctttcttccc catggcccac aacacggct
24721 ccacgcttgcgaa agccatgtc agaaatgaca ccaacgacca gtcctttaat gactacctt
24781 cccggcccaa catgtatccat cccataccgg ccaacgccc caacgtgccc atctccatcc
24841 catcgccaa ctggcagca ttgcgggtt gggcttccac acgcttgaag acaaaggaaa
24901 ccccttccctt gggatcaggc tacgaccctt actacaccta ctctggctcc ataccatacc
24961 ttgacggac cttctatctt aatcacaccc ttaagaagggt ggccattact ttgactctt
25021 ctgttagctg gccggcaac gaccggctgc ttactccca tgagtttggg attaagcgct
25081 cagtttagccgg ggaggctat aacgttagctc agtcaacat gacaaaggac tggttcttag
25141 tgcagatgtt ggccaactac aatattgggtt accagggtt ctacattcca gaaagctaca
25201 aagaccgcatt gtactgtt ttcagaaaact tccaggccat gagccggca gttggggacg
25261 atactaaata caaagatttgc cagcagggtt gaatttccca ccacataac aactcaggct
25321 tcgttaggc cttcgctccc accatgcggc agggacaaggc ttaccccgct aatgttccct
25381 accccactaat aggcaaaaacc gcggttgata gtattaccca gaaaaagttt ctttgcgacc
25441 gcaccctgtg ggcgcatttttgc ttctccagta actttatgtc catgggtgc ctacagacc
25501 tggggccaaaa ctttctctac gcaaaactccg cccacgcgtc agacatgacc ttggaggtgg
25561 atccccatgga cgagccccc ctttctttagt ttttgggatggc aatgttttgc gttggccgtg
25621 tgcaccaggcc gcacccggc gtcatcgaga ccgtgtaccc ggcacgccc ttctcggccg
25681 gcaacgcccac aacataaaaga agcaagcaac atcaacaaca gtcggccca tgggctccag
25741 tgagcaggaa ctgaaagccca ttgtcaaaaga ttttgggtt gggccatatt ttttgggcac
25801 ctatgacaag cgcttcccgag gttttgtttc cccacacaag ctgcctgcg ccatagttaa
25861 cacggccggt cgccgagactg gggggctaca ctggatggcc ttgccttggg acccgccgtc
25921 aaaaacatgc taccttttgc agccctttgg ttttctgtac caacgtctca agcaggttta
25981 ccagtttagg tacgagtccat ttttgcggcgt taggcatttgc ttttgcggcgt
26041 tataacgctg gaaaagtccca cccaaagctg gcaaggcccc aactcggccg cctgtggcc
26101 attctgtctgc atgtttctcc acgcctttgc caactggccc caaaactccca tggatcaca
26161 cccccaccatg aaccttatttgc cgggggtacc caactccatg cttaaacagtccc
26221 gcccaccctg cgccgcacc accggacagctt ctacagctt cttggagcggcc aactcggcc
26281 cttccgcagc cacagtgcgc aatttaggag cgccacttct ttttgcact tggatcaca
26341 gtaaaaaataa tggacttagga gacactttca ataaaggcaaa atgtttttat ttttgcact
26401 tcgggtgatt atttacccca acccttgcggc ttttgcggcgt taaaaatca aagggggttct

FIG. 41

20/92

26461 gccgcgcacat gctatgcgcc actggcaggg acacgttgcg atactggtgt ttatgtctcc
26521 acttaaaactc aggacacaacc atccgcggca gctcggtgaa gtttcaactc cacaggctgc
26581 gcaccatcac caacgcgtt aycaggtcgg ggcggatat ctgaaatcg cagttggggc
26641 ctccgccttg cgcgccgag ttgcataca cagggttaca gcactggaaac actatcagcg
26701 cgggtgggtg cacgctggcc agcagcttct tgcggagat cagatccgcg tccaggctt
26761 cccgttgcg cagggcaac ggagtcaact ttgttagctg cttcccaaaa aagggtgcatt
26821 gcccaggctt tgagttgcac tcgcaccgtt gtggcatcag aagggtgaccg tgcccagtct
26881 gggcgtagg atacagcgc tgcatgaaag cttgtatctg cttaaaagcc acctgagct
26941 ttgcgccttc agagaagaac atgcccgaacttgcggaa aactgttgcg gccggacagg
27001 cccgcgtcatg cacgcacac cttgcgtcgg tggatggat ctgcaccaca ttccggcccc
27061 accgggttcc cagcatcttgc gcttgcgtactg actgtccctt cagcgcgcgc tgcccgttt
27121 cgctcgatc atccatctca atcactgtct ctttattat cataatgttc cctgttagac
27181 acttaagctc gcttcgatc ttagcgcagc ggtgcagcca caacgcgcag cccgtggct
27241 cgtgggtctt gtaggttacc tctgcaaaacg actgcaggta cgcctgcagg aatcgcccc
27301 tcatcgatc aaaggctttt tgctgggtga aggtcagctg caacccgcgg tgctccctgt
27361 tttagccaggctt cttgcatacgc gcccggagag cttccacttgc tgcaggcagt agcttgaagt
27421 ttgccttagt atcgttatcc acgtggtaact tgcgtatcaa cgcgcgcgcga gcctccatgc
27481 ctttctccca cgcagacacg atcggcaggg tcaaggggtt tataccgtt ctttactt
27541 cccgttcaact ggacttcc tttctcttgc gatccgcattt accccgcgcgc actgggtcg
27601 cttcattcag cgcggcacc gtcgcgttac cttcccttgc gtcgttattt agcaccgggt
27661 gggttgcgtaa acccaccatt ttagcgcaca catcttcttgc ttttctctgc tgcgcgcac
27721 tcacctctgg ggatgggggg cgtctgggtt tggaggggggg ggttctttt ttcttttgg
27781 acgcaatggc caaatccgc gtcggatgtcg atggccggg gtcgggttgc cggggcacca
27841 ggcgcatttgc tgacgatcttcttgc ttttgcgttgc cggactcgcg acgcgcgcctc agccgc
27901 tggggggcgc gccccggggc ggcggcgcacg gcaacggggg cggagacgttgc tccatgggt
27961 gttggatgtcg cggccacccg cgtccgcgttgc cgggggttgc ttcgcgttgc tccatgg
28021 gactggccat ttcccttgc tataaggcaga aaaagatcat ggagtcatc gagaaggagg
28081 acagcctaac cggccctttt gaggcgcaca ccacccgcctc caccgatgcc gccaacgcgc
28141 ctaccacccctt cccgcgtcgg gcaacccccc ttaggagggg ggaagtgtatt atcgagcagg
28201 acccaggtt tgcgtatc gacgacgaa gacgacgaaatcgttgc accaacacagag gataaaaagc
28261 aagaccagga cgacgcacag gcaaaacggg aacaagtcgg gccccgggac caaaggcat
28321 ggcgcacttcc agatgtggg gacgacgtgc ttttgcgttgc ttcgcgcgc cagtgc
28381 ttatctgcga cgcgttgc gacgacgtgc atgtggccctt cgcctatgcg gatgtc
28441 ttgcctacgc acgcaccccttgc ttcttgcgttgc gcttgcgttgc ccaacgcggca
28501 catgcgcacccgc caacccgcgc ctcaacttgc accccctattt tgcgttgc accaaccgc
28561 ccacccatca catcttttttgc ctt
28621 gcccggcgga caacgcacgtgc gcttgcgcgc agggcgttgc ctttttttttttttttttt
28681 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc ctttttttttttttttttt
28741 ctctgcataca agaaaacacgc gaaaatggaa gtcactgttgc ttttgcgttgc
28801 gtgacaaacgc ggcgccttgc gtcgttgc gtcacccac ttttgcgttgc
28861 cggcacttac ctt
28921 gtgcacgcacc ctt
28981 ctt
29041 agcgcacgcac gctaatgtat ggcgcacccgc gtcacccac ttttgcgttgc
29101 gtt
29161 gcccgggttgc ctt
29221 ctt
29281 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
29341 ggcacccgcac ctt
29401 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
29461 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
29521 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
29581 ctt
29641 gtgacaaacgc ggcgcacccgc gtcacccac ttttgcgttgc
29701 ctt
29761 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
29821 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
29881 cggcacttac ctttttttttttttttttttttttttttttt
29941 gtgcacgcacc ctttttttttttttttttttttttttttt
29981 ctttttttttttttttttttttttttttttttttttttt
30041 agcgcacgcac gctaatgtat ggcgcacccgc gtcacccac ttttgcgttgc
30101 gtttttttttttttttttttttttttttttttttttttt
30161 gcccgggttgc ctttttttttttttttttttttttttt
30221 ctttttttttttttttttttttttttttttttttttt
30281 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
30341 ggcacccgcac ctttttttttttttttttttttttt
30401 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
30461 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
30521 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
30581 ctttttttttttttttttttttttttttttttttttt
30641 gtgacaaacgc ggcgcacccgc gtcacccac ttttgcgttgc
30701 ctttttttttttttttttttttttttttttttttt
30761 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
30821 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
30881 cggcacttac ctttttttttttttttttttttttt
30941 gtgcacgcacc ctttttttttttttttttttttttt
31001 ctttttttttttttttttttttttttttttttttt
31061 gcccgggttgc ctttttttttttttttttttttt
31121 ctttttttttttttttttttttttttttttttt
31181 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
31241 ggcacccgcac ctttttttttttttttttttttt
31301 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
31361 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
31421 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
31481 ctttttttttttttttttttttttttttttt
31541 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
31601 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
31661 cggcacttac ctttttttttttttttttttttt
31721 gtgcacgcacc ctttttttttttttttttttttt
31781 ctttttttttttttttttttttttttttt
31841 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
31901 ggcacccgcac ctttttttttttttttttttt
31961 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
32021 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
32081 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
32141 ctttttttttttttttttttttttttttt
32201 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
32261 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
32321 cggcacttac ctttttttttttttttttttt
32381 gtgcacgcacc ctttttttttttttttttttt
32441 ctttttttttttttttttttttttttt
32501 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
32561 ggcacccgcac ctttttttttttttttttt
32621 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
32681 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
32741 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
32801 ctttttttttttttttttttttttttt
32861 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
32921 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
32981 cggcacttac ctttttttttttttttttt
33041 gtgcacgcacc ctttttttttttttttttt
33101 ctttttttttttttttttttttttttt
33161 gcccgggttgc ctttttttttttttttt
33221 ctttttttttttttttttttttttt
33281 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
33341 ggcacccgcac ctttttttttttttttt
33401 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
33461 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
33521 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
33581 ctttttttttttttttttttttttt
33641 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
33701 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
33761 cggcacttac ctttttttttttttttt
33821 gtgcacgcacc ctttttttttttttttt
33881 ctttttttttttttttttttttt
33941 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
33981 ggcacccgcac ctttttttttttttttt
34041 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
34101 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
34161 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
34221 ctttttttttttttttttttttttt
34281 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
34341 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
34401 cggcacttac ctttttttttttttttt
34461 gtgcacgcacc ctttttttttttttttt
34521 ctttttttttttttttttttttt
34581 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
34641 ggcacccgcac ctttttttttttttttt
34701 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
34761 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
34821 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
34881 ctttttttttttttttttttttt
34941 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
34981 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
35041 cggcacttac ctttttttttttttttt
35101 gtgcacgcacc ctttttttttttttttt
35161 ctttttttttttttttttttttt
35221 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
35281 ggcacccgcac ctttttttttttttttt
35341 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
35401 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
35461 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
35521 ctttttttttttttttttttttt
35581 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
35641 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
35701 cggcacttac ctttttttttttttttt
35761 gtgcacgcacc ctttttttttttttttt
35821 ctttttttttttttttttttttt
35881 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
35941 ggcacccgcac ctttttttttttttttt
35981 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
36041 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
36101 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
36161 ctttttttttttttttttttttt
36221 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
36281 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
36341 cggcacttac ctttttttttttttttt
36401 gtgcacgcacc ctttttttttttttttt
36461 ctttttttttttttttttttttt
36521 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
36581 ggcacccgcac ctttttttttttttttt
36641 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
36701 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
36761 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
36821 ctttttttttttttttttttttt
36881 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
36941 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
36981 cggcacttac ctttttttttttttttt
37041 gtgcacgcacc ctttttttttttttttt
37101 ctttttttttttttttttttttt
37161 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
37221 ggcacccgcac ctttttttttttttttt
37281 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
37341 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
37401 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
37461 ctttttttttttttttttttttt
37521 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
37581 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
37641 cggcacttac ctttttttttttttttt
37701 gtgcacgcacc ctttttttttttttttt
37761 ctttttttttttttttttttttt
37821 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
37881 ggcacccgcac ctttttttttttttttt
37941 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
37981 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
38041 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
38101 ctttttttttttttttttttttt
38161 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
38221 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
38281 cggcacttac ctttttttttttttttt
38341 gtgcacgcacc ctttttttttttttttt
38401 ctttttttttttttttttttttt
38461 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
38521 ggcacccgcac ctttttttttttttttt
38581 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
38641 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
38701 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
38761 ctttttttttttttttttttttt
38821 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
38881 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
38941 cggcacttac ctttttttttttttttt
38981 gtgcacgcacc ctttttttttttttttt
39041 ctttttttttttttttttttttt
39101 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
39161 ggcacccgcac ctttttttttttttttt
39221 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
39281 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
39341 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
39401 ctttttttttttttttttttttt
39461 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
39521 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
39581 cggcacttac ctttttttttttttttt
39641 gtgcacgcacc ctttttttttttttttt
39701 ctttttttttttttttttttttt
39761 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
39821 ggcacccgcac ctttttttttttttttt
39881 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
39941 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
39981 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
40041 ctttttttttttttttttttttt
40101 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
40161 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
40221 cggcacttac ctttttttttttttttt
40281 gtgcacgcacc ctttttttttttttttt
40341 ctttttttttttttttttttttt
40401 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
40461 ggcacccgcac ctttttttttttttttt
40521 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
40581 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
40641 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
40701 ctttttttttttttttttttttt
40761 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
40821 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
40881 cggcacttac ctttttttttttttttt
40941 gtgcacgcacc ctttttttttttttttt
40981 ctttttttttttttttttttttt
41041 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
41101 ggcacccgcac ctttttttttttttttt
41161 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
41221 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
41281 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
41341 ctttttttttttttttttttttt
41401 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
41461 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
41521 cggcacttac ctttttttttttttttt
41581 gtgcacgcacc ctttttttttttttttt
41641 ctttttttttttttttttttttt
41701 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
41761 ggcacccgcac ctttttttttttttttt
41821 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
41881 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
41941 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
41981 ctttttttttttttttttttttt
42041 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
42101 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
42161 cggcacttac ctttttttttttttttt
42221 gtgcacgcacc ctttttttttttttttt
42281 ctttttttttttttttttttttt
42341 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
42401 ggcacccgcac ctttttttttttttttt
42461 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
42521 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
42581 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
42641 ctttttttttttttttttttttt
42701 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
42761 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
42821 cggcacttac ctttttttttttttttt
42881 gtgcacgcacc ctttttttttttttttt
42941 ctttttttttttttttttttttt
42981 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
43041 ggcacccgcac ctttttttttttttttt
43101 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
43161 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
43221 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
43281 ctttttttttttttttttttttt
43341 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
43401 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
43461 cggcacttac ctttttttttttttttt
43521 gtgcacgcacc ctttttttttttttttt
43581 ctttttttttttttttttttttt
43641 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
43701 ggcacccgcac ctttttttttttttttt
43761 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
43821 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
43881 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
43941 ctttttttttttttttttttttt
43981 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
44041 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
44101 cggcacttac ctttttttttttttttt
44161 gtgcacgcacc ctttttttttttttttt
44221 ctttttttttttttttttttttt
44281 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
44341 ggcacccgcac ctttttttttttttttt
44401 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
44461 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
44521 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
44581 ctttttttttttttttttttttt
44641 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
44701 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
44761 cggcacttac ctttttttttttttttt
44821 gtgcacgcacc ctttttttttttttttt
44881 ctttttttttttttttttttttt
44941 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
44981 ggcacccgcac ctttttttttttttttt
45041 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
45101 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
45161 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
45221 ctttttttttttttttttttttt
45281 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
45341 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
45401 cggcacttac ctttttttttttttttt
45461 gtgcacgcacc ctttttttttttttttt
45521 ctttttttttttttttttttttt
45581 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
45641 ggcacccgcac ctttttttttttttttt
45701 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
45761 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
45821 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
45881 ctttttttttttttttttttttt
45941 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
45981 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
46041 cggcacttac ctttttttttttttttt
46101 gtgcacgcacc ctttttttttttttttt
46161 ctttttttttttttttttttttt
46221 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
46281 ggcacccgcac ctttttttttttttttt
46341 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
46401 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
46461 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
46521 ctttttttttttttttttttttt
46581 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
46641 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
46701 cggcacttac ctttttttttttttttt
46761 gtgcacgcacc ctttttttttttttttt
46821 ctttttttttttttttttttttt
46881 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
46941 ggcacccgcac ctttttttttttttttt
46981 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
47041 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
47101 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
47161 ctttttttttttttttttttttt
47221 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
47281 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
47341 cggcacttac ctttttttttttttttt
47401 gtgcacgcacc ctttttttttttttttt
47461 ctttttttttttttttttttttt
47521 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
47581 ggcacccgcac ctttttttttttttttt
47641 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
47701 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
47761 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
47821 ctttttttttttttttttttttt
47881 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
47941 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
47981 cggcacttac ctttttttttttttttt
48041 gtgcacgcacc ctttttttttttttttt
48101 ctttttttttttttttttttttt
48161 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
48221 ggcacccgcac ctttttttttttttttt
48281 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
48341 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
48401 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
48461 ctttttttttttttttttttttt
48521 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
48581 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
48641 cggcacttac ctttttttttttttttt
48701 gtgcacgcacc ctttttttttttttttt
48761 ctttttttttttttttttttttt
48821 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
48881 ggcacccgcac ctttttttttttttttt
48941 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
48981 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
49041 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
49101 ctttttttttttttttttttttt
49161 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
49221 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
49281 cggcacttac ctttttttttttttttt
49341 gtgcacgcacc ctttttttttttttttt
49401 ctttttttttttttttttttttt
49461 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
49521 ggcacccgcac ctttttttttttttttt
49581 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
49641 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
49701 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
49761 ctttttttttttttttttttttt
49821 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
49881 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
49941 cggcacttac ctttttttttttttttt
49981 gtgcacgcacc ctttttttttttttttt
50041 ctttttttttttttttttttttt
50101 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
50161 ggcacccgcac ctttttttttttttttt
50221 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
50281 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
50341 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
50401 ctttttttttttttttttttttt
50461 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
50521 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
50581 cggcacttac ctttttttttttttttt
50641 gtgcacgcacc ctttttttttttttttt
50701 ctttttttttttttttttttttt
50761 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
50821 ggcacccgcac ctttttttttttttttt
50881 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
50941 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
50981 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
51041 ctttttttttttttttttttttt
51101 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
51161 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
51221 cggcacttac ctttttttttttttttt
51281 gtgcacgcacc ctttttttttttttttt
51341 ctttttttttttttttttttttt
51401 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
51461 ggcacccgcac ctttttttttttttttt
51521 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
51581 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
51641 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
51701 ctttttttttttttttttttttt
51761 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
51821 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
51881 cggcacttac ctttttttttttttttt
51941 gtgcacgcacc ctttttttttttttttt
51981 ctttttttttttttttttttttt
52041 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
52101 ggcacccgcac ctttttttttttttttt
52161 agaactgtctt aacccggggc gtcacccac ttttgcgttgc
52221 cccgcacccgcac ggcgcacccgc gtcacccac ttttgcgttgc
52281 tgcgtatc gccaatccatc ttttgcgttgc ttttgcgttgc
52341 ctttttttttttttttttttttt
52401 ggcacccgcac gtcacccac ttttgcgttgc ttttgcgttgc
52461 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
52521 cggcacttac ctttttttttttttttt
52581 gtgcacgcacc ctttttttttttttttt
52641 ctttttttttttttttttttttt
52701 agggcgaggc ggcgcgcac gtcacccac ttttgcgttgc
52761

FIG. 4J

21/92

29761 gctgcaacct atgcaccccg caccgctccc tggctgcaa ttcacaactg cttagcgaaa
29821 gtcaaattat cggtacctt gagctcagg gtccctcgcc tgacgaaaag tccgcggctc
29881 cggggttgaa actcaactcg gggctgtgga cgtcgctta cttcgcaaa ttgtacctg
29941 aggactacca cgccccacgag attaggtct acgaagacca atcccggcc ccaaagtccgg
30001 agcttaccgc ctgcgtcatt acccaggccc acatccttg ccaattgcaa gccattaaca
30061 aagccgcga agagttctg ctacgaaagg gacgggggt ttacttggac ccccagtccg
30121 gcgaggagct caacccaatc ccccccgcgc cgacggcccta tcagcagccg cgggccttg
30181 cttcccgaga tggcacccaa aagaagctg cagctggccgc cgccgcacc cacggacgag
30241 gaggaaactt gggacagtc ggcagaggag gttttgacg aggaggagga gatgtggaa
30301 gactgggaca gcctagacga ggaagcttcc gaggccgaag aggtgtcaga cgaaacaccg
30361 tcaccctcg tgcattttcc ctcgcggcg ccccaaaaat cggcaaccgt tcccaaggatt
30421 gctacaacct cgcctcctca ggcgcgcgc gcaactcccg ttcgcccacc caaccgtaga
30481 tgggacacca ctggaaaccag ggcggtaag tctaaggcage cgccgcgtt agcccaagag
30541 caacaacagc gccaaggcta cgcgcgttgg cgctgcaca agaacgcac agttgttgc
30601 ttgcaagact gtggggcata catctccctt gcccggcgt ttcttctcta ccatacggc
30661 gtggccttcc cccgttaacat cctgcattac taccgtcata tctacagccc ctactgcacc
30721 ggcggcagcg gcagcaacag cagcggccac gcagaagcaa aggcaaccgg atagcaagac
30781 tctgacaaag cccaaagaaat ccacagcggc ggcagcagca ggaggaggag cactgcgtt
30841 ggcggccaaac gaacccgtat gacccggcga gcttagaaac aggatttttcc caactctgt
30901 tgctatattt caacagagca ggggccaaga acaagagctg aaaataaaaaa acaggctct
30961 ggcgtcccttcc accggcagctt ggcgtatca caaaagcgaat gtcagcttcc ggcgcacgc
31021 ggaagacgcg gaggctcttct ttagcaataa ctgcgcgttgc actttaagg actagtttgc
31081 cgccttttcaaaatctaag cgcggaaaactt acgtcatactc cagcggccac acccggcc
31141 agcacctgtc gtcagcgcctt ttagtgcacaa gggaaatcccc acggccatca tggggaggat
31201 ccagccacaa atgggacttgc cggctggagc tgcccaagac tactcaaccc gaataaacta
31261 catgagcgcg ggaccccaaca ttagatcccg ggtcaacgga atccgcgcacc accgaaaccg
31321 aattctccctc gaacaggcgg accattaccac cacacccgtt aataacctt atccccgttag
31381 ttggcccgct gcccgtgtt accagggaaag tcccgcctcc accactgtgg tactttccag
31441 agacgcccag gccgaagttc agatgactaa ctcagggcgc cagctgcgg gggcgttcc
31501 tcacagggtt cggcgcgcgc ggcagggtat aactcacctg aaaatcagag ggcgaggat
31561 tcagctcaac gacgactcgg ttagcttccctc tcttggtctc cgtccggac ggacattca
31621 gatcgccggc gctggccgctt cttcatattac gcccgtcag gcgatctaa ctctgcagac
31681 ctgcgtccctc gagccgcgcgc cggaggcat tggaaactcta caatttattt aggagttgt
31741 gccttcgggtt tacttcaacc cttttctgg acctccggc cactaccggg accagtttat
31801 tcccaactt gacgcggtaa aagactcggc ggacggttac gactgaatg ccagttggaga
31861 ggcagagcaa ctgcgcctga cacacccctga ccactgcgc cgcacaatgt gctttcccg
31921 cggctccggt gagttttgtt aatttgaattt gcccgaagag catatcgagg gcccggcga
31981 cggcgtccgg ctcaccaccc aggttagact tacacgtac ctgattcggg agtttaccaa
32041 ggcggccctg cttagtggagc gggagcgggg tccctgtgtt ctgaccgtgg tttgaactg
32101 tcctaaccct ggattacatc aagatcttacat tccatcaac taacaataaa cacacaataa
32161 attacttact taaaatcagt cagcaatctt ttgtccagct tattcagcat cacctccctt
32221 ccctcctccc aactctggta ttcagcage ctttagctt cgaactttctt ccaaagtctt
32281 aatgggatgt caaattccctc atgttcttgc ccctccgcac ccactatctt catattgttg
32341 cagatgaaac ggcgcagacc gtctgaagac accttcaacc ctgtgttacc atatgacacg
32401 gaaaccggcc ctccaaactgt gcctttccctt accccctccctt ttgtgtcggc aaatgggtt
32461 caagaaaatgc ccccccggagt gctttcttgc cgtcttcag aacccttggt tacctcacac
32521 ggcgtcttgc cgctaaaaat gggcagcggc ctgtccctgg atcaggcagg caacccttaca
32581 tcaaatacaa tcactgtttc tcaaccgcata aaaaaaacaag tggccacaac ttgccttgc
32641 acatccgcgc cccttacagt cagctcaggc gcccctaacc tggccacaac ttgccttgc
32701 gtggtcttgc acaacactct taccatgcac tcacaaggc cgtctaaccgt gcaagactca
32761 aaacttagea ttgttaccaa agagccactt acgtgttag atggaaaactt gggccctgcag
32821 acatcagccc ccctctctgc cactgtataac aacgcctca tcactactc ctcaccctt
32881 cttactactg caaatggtag tctggcttt accatggaaa aeccaacttta caacaacaaat
32941 gaaaaacttgc ggctcaaaaat tggcggctt cttggacttgc gggccggcactt acatgcacta
33001 acacttagta ctggtcaggg gtttgcaggcataacaatt tgcatacatac aaaagttaca

FIG. 4K

22/92

33061 ggcgcaatag gtttgatac atctggcaac atggaactta aaactggaga tggccttat
 33121 gtggatagcg ccggcctaa caaaaacta catattaatc taaataccac aaaaggcctt
 33181 gctttgaca acaccgcaat aacaattaac gctggaaaag gtttggatt taaaacagac
 33241 tcctcaaactg gaaatcccatt aaaaacaaaat tggatcgatc gcatacaata taataccat
 33301 ggagctatgg ttgcaaaaact tggacagggc ctcagtttgc acagctccgg agccataaca
 33361 atggcagca taaaacaatg cagacttact ctttggacaa caccagaccc atccccaaat
 33421 tgcaatttgg ctcagataa agactgcaag ctaactctgg cgctaaacaaa atgtggcagt
 33481 caaattttgg gcactgtttc agcttggca gtatcaggta atatggcctc catcaatgg
 33541 actctaagca gtgtaaaact gggttcttgc ttttgcatac acggagtgtc tatgtcaat
 33601 tcatcaactgg acaaacagta ttggaaactt agaaacgggg actccactaa cggtaaccca
 33661 tacacttatg ctgttgggtt tatgcacaaac ctaaaagctt accaaaaac taaaagtaaa
 33721 actgcaaaaa gtaatattgt tagccaggta tatcttaatg gtgacaatgc taaaaccattg
 33781 cattttacta ttacgctaaa tggacagat gaaaccaacc aagtaagca atactcaata
 33841 tcattcagtt ggtcctggaa cagttggacaa tacactaatg acaaatttgc caccattcc
 33901 tataccttct cctacattgc ccaggaaataa agaactgtga acctgttgc ttttatgtt
 33961 caacgttggt attttcaat tgcaggaaaat ttcaatgtcat ttttcatca ttagtatagc
 34021 cccaccacca catatcttactatcacc gtaccttaat caaactcaca gaaccctagt
 34081 attcaacctg ccaccccttccccc cccaaacacac agatcaca gtccttctc cccggctggc
 34141 cttaaacacgc atcatatcat gggtaacaga catattctt ggttataat tccacacggt
 34201 ctctgtcga gccaaacgct catcaatgtat gttataaaac tccccggca gctcgcttaa
 34261 gttcatgtcg ctgtccagct gctgagccac aggctgtgtc ccaacttgcgt gtgtctcaac
 34321 gggcggcga ggagaagttc acgcctacat ggggttagag tcataatgtc gcatcaggat
 34381 agggcggtgg tgcgtcagca ggcgcgaat aaactgtgtc cggccggcgt cggctctgc
 34441 ggaataacaac atggcagttgg tctcttcagc gatgattcgc accggccgc gataaggcg
 34501 ccttgcctc cgggcacacg acgcacccct gatctactt aagtcaatgc agtaactgca
 34561 gcacagtacc acaatattgt taaaatccc acatgtcaag ggcgtgtatc caaagctcat
 34621 ggeggggacc acagaaccca cgtggccatc ataccacaag cgcaggtaga ttaagtggcg
 34681 acccctata aacacgttgc acataaaat tacctttt ggcatgttgc aatttcaccac
 34741 ctccggtagc catataaaat tctgattaaa catggcgcca tccaccacca tcctaaacca
 34801 gtcggccaaa acctggccgc cggctatgc ctgcaggaa cgggactgg aacaatgaca
 34861 gtggagagcc caggactgt aaccatggat catcatgtc gtcatgatcat caatgttggc
 34921 acaacacagg cacacgtca tacacttccct caggattaca agtctctcc ggcgtcagaac
 34981 catatcccg ggaacaaccc attctgaat cagcttaaat cccacactgc agggaaagacc
 35041 tcgcacgtaa ctcacgttgc gcatgtcaat agtggatcat tcggcggca gggatgatc
 35101 ctccagttatg gtagcgggg tttctgtctc aaaaggaggt agacgttcc tactgtacgg
 35161 agtgcggca gacaaccgg atcgttgg tgcgtatgtc atgccaatgc gaaacggcga
 35221 cgtagtcata tttctgttgc aaaaaccgg tgcggcgtg aaaaacagat ctgcgtctcc
 35281 ggtctcgccg ctttagatgc tctgttgc tttgttgc ttttgcataat ttttgcataat
 35341 ccaggcgccc cctggctcg ggttctatg aaactccttc atgcggcgct gcccgtataa
 35401 catccaccac cgcagaataa gcccacccca gccaacccatc acattcgatc tgcaatgc
 35461 acacggggagg agcgggaaaga gtcggaaagaa ccatgtttt ttttttattc caaaagat
 35521 tccaaacccct caaaatggat atcttataat ttttgcataat ttttgcataat
 35581 aactctacag cccaaacccata gataatggca tttgttgc ttttgcataat ggcgttccaa
 35641 agccaaacgg ccctcactgc caagtggacg taaaaggctaa acccttcagg gtgaatctcc
 35701 tctataaaaca ttccagcacc ttcaaccatg cccaaataat ttttgcataat ttttgcataat
 35761 aatataatctc taagcaatc ccgaatatta agtccggcca ttgtaaaaat ctgtccaga
 35821 ggcgcctcca ctttcagcc caagcagcgatc atcatgttgc ctttgcataat ggttccatc
 35881 agacgttat aagattcaaa agcggaaat taacaaaaat acccttcagg gtgaatctcc
 35941 ttccggcggc cagctgaaata taatgtca ggtctcgacg gaccggcgcc gccacttcc
 36001 cggccaggaaac catgacaaaaa gaaacccacac tgattatgac acgcataactc ggagctatgc
 36061 taaccagcgt agcccccgt ttttgcataat ttttgcataat ggttccatc
 36121 tgctcaaaaaa atcaggcaaa gcctcgccaa aaaaagaaaaag cccatcgatc ttttgcataat
 36181 gcagataaaag gcaggtaagc tccggaaacca ccacagaaaaa agacaccatt ttttgcataat
 36241 acatgtctgc gggtttctgc ataaacacaa aaaaaataa caaaaaaaca tttaaacatt
 36301 agaaggctgt cttacaacag gaaaaacaaac ctttataacg ataaagacggc atacggccat

FIG. 4L

23/92

36361 gccggcgtga ccgtaaaaaaaaa actggtcacc gtgattaaaaa agcaccaccc agagctccctc
36421 ggtcatgtcc ggagtctataa tgtaagactc gtaaaacaca tcaggttgat tcacatcggt
36481 cagtctaaa aagcgaccga aatagcccg gggatacat acccgaggc gtagagacaa
36541 cattacagcc cccataggag gtataacaaa attaatagga gaaaaaaca cataaacacc
36601 tgaaaaaccc tcctgcctag gaaaaatagc accctccgc tccagaacaa catacagcgc
36661 ttccacagcg gcagccataa cagtcagcct taccagtaaa aaagaaaaacc tattaaaaaa
36721 acaccactcg acacggcacc agctaatca gtcacagtgt aaaaaaggc caagtgcaga
36781 gcgagtataat ataggactaa aaaatgacgt aacggtaaa gtccacaaaa aacacccaga
36841 aaaccgcacg cgaacctacg cccagaaaacg aaagccaaaa aaccacaaac ttccctcaaatt
36901 cgtcaattcc gttttccac gttacgtcac ttcccatttt aagaaaaacta caattcccaa
36961 cacatacaag ttactccgcc ctaaaaccta cgtcacccgc cccgtcccc cggcccgcc
37021 cacgtcacaa actccacccc ctcattatca tattggcttc aatccaaaaat aaggtatatt
37081 attgatgatg

FIG. 4M

24/92

10	30	50
ATGGCGCCCATCACGGCCTACTCCAACAGACGCGGGGCTACTTGTTGCATCATCACT		
-----+-----+-----+-----+-----+-----+		
MetAlaProIleThrAlaTyrSerGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr		
10		20
70	90	110
AGCCTTACAGGCCGGACAAGAACCAAGGTCGAGGGAGAGGTTCAGGTGGTTCCACCGCA		
-----+-----+-----+-----+-----+-----+		
SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnValValSerThrAla		
30		40
130	150	170
ACACAATCCTCCTGGCGACCTGCGTCAACGGCGTGTGGACCGTTACCATGGTGCT		
-----+-----+-----+-----+-----+-----+		
ThrGlnSerPheLeuAlaThrCysValAsnGlyValCysTrpThrValTyrHisGlyAla		
50		60
190	210	230
GGCTCAAAGACCTTAGCCGGCCAAAGGGGCAATCACCCAGATGTACACTAATGTGGAC		
-----+-----+-----+-----+-----+-----+		
GlySerLysThrLeuAlaGlyProLysGlyProIleThrGlnMetTyrThrAsnValAsp		
70		80
250	270	290
CAGGACCTCGTCGGCTGGCAGGGCGCCCCCGGGCGCGTTCCCTGACACCATGCACCTGT		
-----+-----+-----+-----+-----+-----+		
GlnAspLeuValGlyTrpGlnAlaProProGlyAlaArgSerLeuThrProCysThrCys		
90		100
310	330	350
GGCAGCTCAGACCTTACTTGGTCACGAGACATGCTGACGTCAATTCCGGTGCGCCGGCG		
-----+-----+-----+-----+-----+-----+		
GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg		
110		120
370	390	410
GGCGACAGTAGGGGGAGCCTGCTCTCCCCCAGGCCTGTCCTACTTGAAGGGCTCTTCG		
-----+-----+-----+-----+-----+-----+		
GlyAspSerArgGlySerLeuLeuSerProArgProValSerTyrLeuLysGlySerSer		
130		140

FIG. 5A

25/92

430	450	470
GGTGGTCCACTGCTCTGCCCTTCGGGGCACGCTGTGGGCATCTCCGGGTGCCGTATGC		
-----+-----+-----+-----+-----+-----+		
GlyGlyProLeuLeuCysProSerGlyHisAlaValGlyIlePheArgAlaAlaValCys		
150		160
-----+-----+-----+-----+-----+-----+		
490	510	530
ACCCGGGGGGTTGCGAAGGCGGTGGACTTTGTGCCCGTACAGTCCATGGAAACTACTATG		
-----+-----+-----+-----+-----+-----+		
ThrArgGlyValAlaLysAlaValAspPheValProValGluSerMetGluThrThrMet		
170		180
-----+-----+-----+-----+-----+-----+		
550	570	590
CGGTCTCCGGTCTCACGGACAACATCCCCCCCCGGCGTACCGCAGTCATTCAAGTG		
-----+-----+-----+-----+-----+-----+		
ArgSerProValPheThrAspAsnSerSerProProAlaValProGlnSerPheGlnVal		
190		200
-----+-----+-----+-----+-----+-----+		
610	630	650
GCCCACCTACACGCTCCACTGGCAGCGGCAAGAGTACTAAAGTGCCGGCTGCATATGCA		
-----+-----+-----+-----+-----+-----+		
AlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAla		
210		220
-----+-----+-----+-----+-----+-----+		
670	690	710
GCCCAAGGGTACAAGGTGCTCGCTCAATCCGTCCGTTGCCGCTACCTTAGGGTTGGG		
-----+-----+-----+-----+-----+-----+		
AlaGinGlyTyrLysValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGly		
230		240
-----+-----+-----+-----+-----+-----+		
730	750	770
GCGTATATGTCTAAGGCACACGGTATTGACCCCCAACATCAGAACTGGGTAAGGACCATT		
-----+-----+-----+-----+-----+-----+		
AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle		
250		260
-----+-----+-----+-----+-----+-----+		
790	810	830
ACCAACAGGCGCCCCCGTCACATACTCTACCTATGGCAAGTTCTGCCGATGGTGGTTGC		
-----+-----+-----+-----+-----+-----+		
ThrThrGlyAlaProValThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys		
270		280

FIG. 5B

26/92

850	870	890
TCTGGGGCGCTTATGACATCATAATATGTGATGAGTGCCATTCAACTGACTCGACTACA		
-----+-----+-----+-----+-----+		
SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspSerThrThr		
290		
300		
910	930	950
ATCTTGGGCATCGGCACAGTCCTGGACCAAGCGGAGACGGCTGGAGCGCGGCTTGTCTG		
-----+-----+-----+-----+-----+		
IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal		
310		
320		
970	990	1010
CTCGCCACCGCTACGCCCTCCGGGATCGGTACCCGTGCCACACCCAAACATCGAGGAGGTG		
-----+-----+-----+-----+-----+		
LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal		
330		
340		
1030	1050	1070
GCCCTGTCTAATACTGGAGAGATCCCCCTCTATGGCAAAGCCATCCCCATTGAAGCCATC		
-----+-----+-----+-----+-----+		
AlaLeuSerAsnThrGlyGluIleProPheTyrGlyLysAlaIleProIleGluAlaIle		
350		
360		
1090	1110	1130
AGGGGGGGAGGCATCTCATTTCTGTCATTCCAAGAAGAAGTGGGACGAGCTCGCCGCA		
-----+-----+-----+-----+-----+		
ArgGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAla		
370		
380		
1150	1170	1190
AAGCTGTCAGGCCTCGGAATCAACGCTGTGGCGTATTACCGGGGCTCGATGTGTCCGTC		
-----+-----+-----+-----+-----+		
LysLeuSerGlyLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal		
390		
400		
1210	1230	1250
ATACCAAATCGGAGACGTCGTTGTCGTGGCAACAGACGCTCTGATGACGGGCTATACG		
-----+-----+-----+-----+-----+		
IleProThrIleGlyAspValValValValAlaThrAspAlaLeuMetThrGlyTyrThr		
410		
420		

FIG. 5C

27/92

1270	1290	1310
GGCGACTTGACTCAGTGATCGACTGTAACACATGTGTCACCCAGACAGTCGACTTCAGC		
-----+-----+-----+-----+-----+-----+		
GlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrValAspPheSer		
430		440
-----+-----+-----+-----+-----+-----+		
1330	1350	1370
TTGGATCCCACCTTCACCATTGAGACGACGACCGTGCCTCAAGACGCAGTGTGCGCTCG		
-----+-----+-----+-----+-----+-----+		
LeuAspProThrPheThrIleGluThrThrValProGlnAspAlaValSerArgSer		
450		460
-----+-----+-----+-----+-----+-----+		
1390	1410	1430
CAGCGGCGGGTAGGACTGGCAGGGTAGGAGAGGCATCTACAGGTTGTGACTCCGGGA		
-----+-----+-----+-----+-----+-----+		
GlnArgArgGlyArgThrGlyArgGlyArgGlyIleTyrArgPheValThrProGly		
470		480
-----+-----+-----+-----+-----+-----+		
1450	1470	1490
GAACGGCCCTCGGGCATGTTGATTCCCTCGTCCCTGTGAGTGCTATGACGCGGGCTGT		
-----+-----+-----+-----+-----+-----+		
GluArgProSerGlyMetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCys		
490		500
-----+-----+-----+-----+-----+-----+		
1510	1530	1550
GCTTGGTACGAGCTCACCCCCGCGAGACCTCGGTTAGGTTGCGGGCCTACCTGAACACA		
-----+-----+-----+-----+-----+-----+		
AlaTrpTyrGluLeuThrProAlaGluThrSerValArgLeuArgAlaTyrLeuAsnThr		
510		520
-----+-----+-----+-----+-----+-----+		
1570	1590	1610
CCAGGGTTGCCCGTTGCCAGGACCACCTGGAGTTCTGGGAGAGTGTCTTCACAGGCCCTC		
-----+-----+-----+-----+-----+-----+		
ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPheThrGlyLeu		
530		540
-----+-----+-----+-----+-----+-----+		
1630	1650	1670
ACCCACATAGATGCACACTTCTTGCCAGACCAAGCAGGAGACAACCTCCCTAC		
-----+-----+-----+-----+-----+-----+		
ThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsnPheProTyr		
550		560

FIG. 5D

28/92

1690	1710	1730
CTGGTAGCATACCAAGCCACGGTGTGCGCCAGGGCTCAGGCCCCACCTCCATCATGGGAT		
-----+-----+-----+-----+-----+-----+		
LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAsp		
570		
580		
1750	1770	1790
CAAATGTGGAAGTGTCTCATACGGCTGAAACCTACGCTGCACGGGCCAACACCCCTTGCTG		
-----+-----+-----+-----+-----+-----+		
GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu		
590		
600		
1810	1830	1850
TACAGGCTGGGAGCCGTCCAAAATGAGGTACCCCTCACCCACCCATAACCAAATACATC		
-----+-----+-----+-----+-----+-----+		
TyrArgLeuGlyAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIle		
610		
620		
1870	1890	1910
ATGGCATGCATGTCGGCTGACCTGGAGGTGTCACTAGCACCTGGGTGCTGGTGGCGGA		
-----+-----+-----+-----+-----+-----+		
MetAlaCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly		
630		
640		
1930	1950	1970
GTCCTTGCAGCTCTGGCCGCGTATTGCCTGACAACAGGCAGTGTGGTCATTGTGGTAGG		
-----+-----+-----+-----+-----+-----+		
ValLeuAlaAlaLeuAlaAlaTyrCysLeuThrThrGlySerValValIleValGlyArg		
650		
660		
1990	2010	2030
ATTATCTTGTCCGGGAGGCCGGCTATTGTTCCCGACAGGGAGTTCTCTACCAGGAGTTC		
-----+-----+-----+-----+-----+-----+		
IleIleLeuSerGlyArgProAlaIleValProAspArgGluPheLeuTyrGlnGluPhe		
670		
680		
2050	2070	2090
GATGAAATGGAAGAGTGCACCTCGCACCTCCCTACATCGAGCAGGGATGCAGCTGCC		
-----+-----+-----+-----+-----+-----+		
AspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGlnGlyMetGlnLeuAla		
690		
700		

FIG. 5E

29/92

2110	2130	2150
GAGCAATTCAAGCAGAAAGCGCTGGGTTACTGCAAACAGCCACCAACAAGCGGAGGCT		
-----+-----+-----+-----+-----+-----+		
GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaThrLysGlnAlaGluAla		
710		720
2170	2190	2210
GCTGCTCCGTGGTGGAGTCCAAGTGGCGAGCCCTTGAGACATTCTGGCGAACATG		
-----+-----+-----+-----+-----+-----+		
AlaAlaProValValGluSerLysTrpArgAlaLeuGluThrPheTrpAlaLysHisMet		
730		740
2230	2250	2270
TGGAATTTCATCAGCGGGATACAGTACTTAGCAGGCTTATCCACTCTGCCTGGGAACCCC		
-----+-----+-----+-----+-----+-----+		
TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro		
750		760
2290	2310	2330
GCAATAGCATCATTGATGGCATTACAGCCTCTATCACCAAGCCCCGCTCACCAACCCAAAGT		
-----+-----+-----+-----+-----+-----+		
AlaIleAlaSerLeuMetAlaPheThrAlaSerIleThrSerProLeuThrThrGlnSer		
770		780
2350	2370	2390
ACCCCTCTGTTAACATCTGGGGGGTGGCTGCCCAACTGCCCCCCCCCAGCGCC		
-----+-----+-----+-----+-----+-----+		
ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaProProSerAla		
790		800
2410	2430	2450
GCTTCGGCTTCGTGGCGCCGGCATGCCGGTGGCTGTTGGCAGCATAGGCCTTGGG		
-----+-----+-----+-----+-----+-----+		
AlaSerAlaPheValGlyAlaGlyIleAlaGlyAlaAlaValGlySerIleGlyLeuGly		
810		820
2470	2490	2510
AAGGTGCTGTGGACATTCTGGCGGGTATGGAGCAGGAGTGGCCGGCGCGCTCGTGGCC		
-----+-----+-----+-----+-----+-----+		
LysValLeuValAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla		
830		840

FIG. 5F

30/92

2530	2550	2570
TTCAAGGTATGAGCGGCCGAGATGCCCTCCACCGAGGACCTGGTCAATCTACTTCCTGCC		
-----+-----+-----+-----+-----+-----+		
PheLysValMetSerGlyGluMetProSerThrGluAspLeuValAsnLeuLeuProAla		
850		860
2590	2610	2630
ATCCTCTCTCCTGGCGCCCTGGTCGTGGGTCGTGTGCAGCAATACTGCGTCGACAC		
-----+-----+-----+-----+-----+-----+		
IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis		
870		880
2650	2670	2690
GTGGGTCCGGGAGAGGGGGCTGTGCAGTGGATGAACCGGCTGATAGCGTTCGCCTCGCG		
-----+-----+-----+-----+-----+-----+		
ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg		
890		900
2710	2730	2750
GGTAATCATGTTTCCCCCACGCACTATGTGCCTGAGAGCGACGCCGAGCGCTGTTACT		
-----+-----+-----+-----+-----+-----+		
GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr		
910		920
2770	2790	2810
CAGATCCTCTCCAGCCTTACCATCACTCAGCTGCTGAAAAGGCTCACCAGTGGATTAAT		
-----+-----+-----+-----+-----+-----+		
GlnIleLeuSerSerLeuThrIleThrGlnLeuLeuLysArgLeuHisGlnTrpIleAsn		
930		940
2830	2850	2870
GAAGACTGCTCCACACCGTGTTCGGCTCGGCTAAGGGATGTTGGGACTGGATATGC		
-----+-----+-----+-----+-----+-----+		
GluAspCysSerThrProCysSerGlySerTrpLeuArgAspValTrpAspTrpIleCys		
950		960
2890	2910	2930
ACGGTGTGACTGACTTCAAGACCTGGCTCCAGTCCAAGCTCCTGCCGAGCTACCGGGA		
-----+-----+-----+-----+-----+-----+		
ThrValLeuThrAspPheLysThrTrpLeuGlnSerLysLeuLeuProGlnLeuProGly		
970		980

FIG. 5G

31/92

2950	2970	2990
GTCCCTTTTCTCGCCAACGCGGGTACAAGGGAGTCTGGCGGGAGACGGCATCATG		
-----+-----+-----+-----+-----+		
ValProPhePheSerCysGlnArgGlyTyrLysGlyValTrpArgGlyAspGlyIleMet		
990		
1000		
3010	3030	3050
CAAACCACTGCCATGTGGAGCACAGATCACCGGACATGTCAAAACGGTCCATGAGG		
-----+-----+-----+-----+-----+		
GlnThrThrCysProCysGlyAlaGlnIleThrGlyHisValLysAsnGlySerMetArg		
1010		
1020		
3070	3090	3110
ATCGTCGGGCTAAGACCTGCAGAACACGTGGCATGGAACATTCCCCATCAACGCATAC		
-----+-----+-----+-----+-----+		
IleValGlyProLysThrCysSerAsnThrTrpHisGlyThrPheProIleAsnAlaTyr		
1030		
1040		
3130	3150	3170
ACCACGGGCCCCCTGCACACCCCTCTCCAGCGCCAAACTATTCTAGGGCGCTGTGGCGGGTG		
-----+-----+-----+-----+-----+		
ThrThrGlyProCysThrProSerProAlaProAsnTyrSerArgAlaLeuTrpArgVal		
1050		
1060		
3190	3210	3230
GCCGCTGAGGAGTACGTGGAGGTACGCCAGGTTCCAGCGGGTGGGGATTCACACTACGTGACGGGCATG		
-----+-----+-----+-----+-----+		
AlaAlaGluGluTyrValGluValThrArgValGlyAspPheHisTyrValThrGlyMet		
1070		
1080		
3250	3270	3290
ACCACTGACAACGTAAAGTCCCCATGCCAGGTTCCGGCTCCTGAATTCTTCACGGAGGTG		
-----+-----+-----+-----+-----+		
ThrThrAspAsnValLysCysProCysGlnValProAlaProGluPhePheThrGluVal		
1090		
1100		
3310	3330	3350
GACGGAGTGCAGGTTGCACAGGTACGTCCGGCGTGCAGGCCTCTCCTACGGGAGGAGGTT		
-----+-----+-----+-----+-----+		
AspGlyValArgLeuHisArgTyrAlaProAlaCysArgProLeuLeuArgGluGluVal		
1110		
1120		

FIG. 5H

32/92

3370	3390	3410
ACATTCCAGGTCTGGCTCAACCAATACCTGGTTGGTCACAGCTACCATGCGAGCCCCGAA		
-----+-----+-----+-----+-----+-----+		
ThrPheGlnValGlyLeuAsnGlnTyrLeuValGlySerGlnLeuProCysGluProGlu		
1130		1140
-----+-----+-----+-----+-----+-----+		
3430	3450	3470
CCGGATGTAGCAGTGCTCACTTCCATGCTACCGACCCCTCCCACATCACAGCAGAAACG		
-----+-----+-----+-----+-----+-----+		
ProAspValAlaValLeuThrSerMetLeuThrAspProSerHisIleThrAlaGluThr		
1150		1160
-----+-----+-----+-----+-----+-----+		
3490	3510	3530
GCTAAGCGTAGGTTGGCCAGGGGGTCTCCCCCCTCCTGGCCAGCTTCAGCTAGCCAG		
-----+-----+-----+-----+-----+-----+		
AlaLysArgArgLeuAlaArgGlySerProProSerLeuAlaSerSerSerAlaSerGln		
1170		1180
-----+-----+-----+-----+-----+-----+		
3550	3570	3590
TTGTCTGCGCTTCCTTGAGGGCAGATGCACTACCCACCATGTCCTCCGGACGCTGAC		
-----+-----+-----+-----+-----+-----+		
LeuSerAlaProSerLeuLysAlaThrCysThrThrHisHisValSerProAspAlaAsp		
1190		1200
-----+-----+-----+-----+-----+-----+		
3610	3630	3650
CTCATCGAGGCCAACCTCCTGTGGCGGCAGGAGATGGGGGGAACATCACCCGGTGGAG		
-----+-----+-----+-----+-----+-----+		
LeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGlu		
1210		1220
-----+-----+-----+-----+-----+-----+		
3670	3690	3710
TCGGAGAACAAAGGTGGTAGTCCTGGACTCTTCGACCCGTTCGAGCGGAGGAGGATGAG		
-----+-----+-----+-----+-----+-----+		
SerGluAsnLysValValValLeuAspSerPheAspProLeuArgAlaGluGluAspGlu		
1230		1240
-----+-----+-----+-----+-----+-----+		
3730	3750	3770
AGGGAAAGTATCCGTCCGGCGGAGATCCTGCGGAAATCCAAGAAGTTCCCCGAGCGATG		
-----+-----+-----+-----+-----+-----+-----+		
ArgGluValSerValProAlaGluIleLeuArgLysSerLysLysPheProAlaAlaMet		
1250		1260

FIG. 5I

33/92

3790	3810	3830
CCCATCTGGCGCGCCCGGATTACAACCCCTCCACTGTTAGAGTCCTGGAAGGACCCGGAC		
-----+-----+-----+-----+-----+-----+		
ProIleTrpAlaArgProAspTyrAsnProProLeuLeuGluSerTrpLysAspProAsp		
1270		1280
3850	3870	3890
TACGTCCCTCCGGTGGTGCACGGGTGCCCGTTGCCACCTATCAAGGCCCCCTCCAATACCA		
-----+-----+-----+-----+-----+-----+		
TyrValProProValValHisGlyCysProLeuProProIleLysAlaProProIlePro		
1290		1300
3910	3930	3950
CCTCCACGGAGAAAGAGGACGGTTGTCTAACAGAGTCCTCCGTGTCTCTGCCTTAGCG		
-----+-----+-----+-----+-----+-----+		
ProProArgArgLysArgThrValValLeuThrGluSerSerValSerSerAlaLeuAla		
1310		1320
3970	3990	4010
GAGCTCGCTACTAACACCTTCGGCAGCTCCGAATCATCGGCCGTCGACAGCGGCACGGCG		
-----+-----+-----+-----+-----+-----+		
GluLeuAlaThrLysThrPheGlySerSerGluSerSerAlaValAspSerGlyThrAla		
1330		1340
4030	4050	4070
ACCGCCCTTCCTGACCAGGCCCTCGACGACGGTGACAAAGGATCCGACGTTGAGTCGTAC		
-----+-----+-----+-----+-----+-----+		
ThrAlaLeuProAspGlnAlaSerAspAspGlyAspLysGlySerAspValGluSerTyr		
1350		1360
4090	4110	4130
TCCTCCATGCCCCCCCCTTGAGGGGGAACCGGGGGACCCCGATCTCAGTGACGGGTCTTGG		
-----+-----+-----+-----+-----+-----+		
SerSerMetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrp		
1370		1380
4150	4170	4190
TCTACCGTGAGCGAGGAAGCTAGTGAGGATGTCTGCTGCTCAATGTCCTACACATGG		
-----+-----+-----+-----+-----+-----+		
SerThrValSerGluGluAlaSerGluAspValValCysCysSerMetSerTyrThrTrp		
1390		1400

FIG. 5J

34/92

4210	4230	4250
ACAGGCGCCTTGATCACGCCATGGCTGCGAGGAAAGCAAGCTGCCATCAACGCGTTG		
-----+-----+-----+-----+-----+		
ThrGlyAlaLeuIleThrProCysAlaAlaGluGluSerLysLeuProIleAsnAlaLeu		
1410		1420
4270	4290	4310
AGCAACTTTGCTGCGCCACCATAACATGTTTATGCCACAACATCTCGCAGCGCAGGC		
-----+-----+-----+-----+-----+		
SerAsnSerLeuLeuArgHisHisAsnMetValTyrAlaThrThrSerArgSerAlaGly		
1430		1440
4330	4350	4370
CTGCGGCAGAAGAAGGTACCTTGACAGACTGCAAGTCCTGGACGACCACTACCGGGAC		
-----+-----+-----+-----+-----+		
LeuArgGlnLysLysValThrPheAspArgLeuGlnValLeuAspAspHisTyrArgAsp		
1450		1460
4390	4410	4430
GTGCTCAAGGAGATGAAGCGAAGGCGTCCACAGTTAAGGCTAAACTCCTATCCGTAGAG		
-----+-----+-----+-----+-----+		
ValLeuLysGluMetLysAlaLysAlaSerThrValLysAlaLysLeuLeuSerValGlu		
1470		1480
4450	4470	4490
GAAGCCTGCAAGCTGACGCCACATTGGCAAATCCAAGTTGGCTATGGGCAG		
-----+-----+-----+-----+-----+		
GluAlaCysLysLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLys		
1490		1500
4510	4530	4550
GACGTCCGGAACCTATCCAGCAAGGCCGTTAACCAACATCCACTCCGTGTGGAAGGACTTG		
-----+-----+-----+-----+-----+		
AspValArgAsnLeuSerSerLysAlaValAsnHisIleHisSerValTrpLysAspLeu		
1510		1520
4570	4590	4610
CTGGAAGACACTGTGACACCAATTGACACCCACCATGGCAAAAATGAGGTTTCTGT		
-----+-----+-----+-----+-----+		
LeuGluAspThrValThrProIleAspThrIleMetAlaLysAsnGluValPheCys		
1530		1540

FIG. 5K

35/92

4630	4650	4670
GTCCAACCAGAGAAAGGAGGCCGTAAGCCAGCCCCGCTTATCGTATTCCCAGATCTGGGA		
-----+-----+-----+-----+-----+-----+		
ValGlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGly		
1550		1560
-----+-----+-----+-----+-----+-----+		
4690	4710	4730
GTCCTGTATGCGAGAAGATGGCCCTCTATGATGTGGCTCCACCCCTCAGGTCGTG		
-----+-----+-----+-----+-----+-----+		
ValArgValCysGluLysMetAlaLeuTyrAspValValSerThrLeuProGlnValVal		
1570		1580
-----+-----+-----+-----+-----+-----+		
4750	4770	4790
ATGGGCTCCTCATACGGATTCCAGTACTCTCCTGGGCAGCGAGTCGAGTTCCCTGGTGAAT		
-----+-----+-----+-----+-----+-----+		
MetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValAsn		
1590		1600
-----+-----+-----+-----+-----+-----+		
4810	4830	4850
ACCTGGAAATCAAAGAAAAACCCCATGGGCTTTCATATGACACTCGCTGTTGACTCA		
-----+-----+-----+-----+-----+-----+		
ThrTrpLysSerLysLysAsnProMetGlyPheSerTyrAspThrArgCysPheAspSer		
1610		1620
-----+-----+-----+-----+-----+-----+		
4870	4890	4910
ACGGTCACCGAGAACGACATCCGTGGAGGAGTCATTTACCAATGTTGTGACTTGGCC		
-----+-----+-----+-----+-----+-----+		
ThrValThrGluAsnAspIleArgValGluGluSerIleTyrGlnCysCysAspLeuAla		
1630		1640
-----+-----+-----+-----+-----+-----+		
4930	4950	4970
CCCCGAAGCCAGACAGGCCATAAAATCGCTCACAGAGGGCTTTATATCGGGGGTCCCTCTG		
-----+-----+-----+-----+-----+-----+		
ProGluAlaArgGlnAlaIleLysSerLeuThrGluArgLeuTyrIleGlyGlyProLeu		
1650		1660
-----+-----+-----+-----+-----+-----+		
4990	5010	5030
ACTAATTCAAAAGGGCAGAACTGGGTTATGCCGGTGCCGCGCAGCGGGCGTGCTGACG		
-----+-----+-----+-----+-----+-----+		
ThrAsnSerLysGlyGlnAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThr		
1670		1680

FIG. 5L

36/92

5050	5070	5090
ACTAGCTGCGGTAACACCCCTCACATGTTACTTGAAGGCCTCTGCAGCCTGTCGAGCTGCG		
-----+-----+-----+-----+-----+-----+		
ThrSerCysGlyAsnThrLeuThrCysTyrLeuLysAlaSerAlaAlaCysArgAlaAla		
1690		
1700		
5110	5130	5150
AAGCTCCAGGACTGCACGATGCTCGTGAAACGGAGACGACCTTGTGTTATCTGTGAAAGC		
-----+-----+-----+-----+-----+-----+		
LysLeuGlnAspCysThrMetLeuValAsnGlyAspAspLeuValValIleCysGluSer		
1710		
1720		
5170	5190	5210
GCGGGAAACCCAAAGAGGGACGCCGGCGAGCCTACGAGTCTTCACGGAGGCTATGACTAGGTAC		
-----+-----+-----+-----+-----+-----+		
AlaGlyThrGlnGluAspAlaAlaSerLeuArgValPheThrGluAlaMetThrArgTyr		
1730		
1740		
5230	5250	5270
TCTGCCCCCCCCGGGGACCCGCCAACAGAAATACGACTTGGAGCTGATAACATCATGT		
-----+-----+-----+-----+-----+-----+		
SerAlaProProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCys		
1750		
1760		
5290	5310	5330
TCCTCCAATGTGTGGTCGCCACGATGCATCAGGCAAAGGGTGTACTACCTCACCGT		
-----+-----+-----+-----+-----+-----+		
SerSerAsnValSerValAlaHisAspAlaSerGlyLysArgValTyrTyrLeuThrArg		
1770		
1780		
5350	5370	5390
GATCCCCACCACCCCCCTCGCACGGGCTGCGTGGAAACAGCTAGACACACTCCAGTTAAC		
-----+-----+-----+-----+-----+-----+		
AspProThrThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsn		
1790		
1800		
5410	5430	5450
TCCTGGCTAGGCAACATTATCATGTATGCCCACTTGTGGGCAAGGATGATTCTGATG		
-----+-----+-----+-----+-----+-----+		
SerTrpLeuGlyAsnIleIleMetTyrAlaProThrLeuTrpAlaArgMetIleLeuMet		
1810		
1820		

FIG. 5M

37/92

5470

5490

5510

ACTCACTCTTCTCCATCCTCTAGCACAGGAGCAACTTGAAAAAGCCCTGGACTGCCAG
 -----+-----+-----+-----+-----+-----+
 ThrHisPhePheSerIleLeuLeuAlaGlnGluGlnLeuGluLysAlaLeuAspCysGln
 1830 1840

5530

5550

5570

ATCTACGGGCCTGTTACTCCATTGAGCCACTTGACCTACCTCAGATCATTGAACGACTC
 -----+-----+-----+-----+-----+
 IleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuProGlnIleIleGluArgLeu
 1850 1860

5590

5610

5630

CATGGCCTTAGCGCATTTCACTCCATAGTTACTCTCCAGGTGAGATCAATAGGGTGGCT
 -----+-----+-----+-----+-----+
 HisGlyLeuSerAlaPheSerLeuHisSerTyrSerProGlyGluIleAsnArgValAla
 1870 1880

5650

5670

5690

TCATGCCTCAGGAAACTTGGGGTACCAACCCCTTGCAGTCTGGAGACATCGGCCAGGAGC
 -----+-----+-----+-----+-----+
 SerCysLeuArgLysLeuGlyValProProLeuArgValTrpArgHisArgAlaArgSer
 1890 1900

5710

5730

5750

GTCCCGCTAGGCTACTGTCCCAGGGGGGAGGGCCGCCACTTGTGGCAAGTACCTCTTC
 -----+-----+-----+-----+-----+
 ValArgAlaArgLeuLeuSerGlnGlyGlyArgAlaAlaThrCysGlyLysTyrLeuPhe
 1910 1920

5770

5790

5810

AACTGGGCAGTGAAGACCAAACCTCAACTCACTCCAATCCCGGCTGCGTCCCAGCTGGAC
 -----+-----+-----+-----+-----+
 AsnTrpAlaValLysThrLysLeuLysLeuThrProIleProAlaAlaSerGlnLeuAsp
 1930 1940

5830

5850

5870

TTGTCCGGCTGGTTCGTTGCTGGTTACAGGGGGGAGACATATATCACAGCCTGTCTCGT
 -----+-----+-----+-----+-----+
 LeuSerGlyTrpPheValAlaGlyTyrSerGlyGlyAspIleTyrHisSerLeuSerArg
 1950 1960

FIG. 5N

38/92

5890

5910

5930

GCCCCGACCCCGCTGGTTCATGCTGTGCCTACTCCTACTTTCTGTAGGGGTAGGCATCTAC
-----+-----+-----+-----+-----+-----+
AlaArgProArgTrpPheMetLeuCysLeuLeuLeuLeuSerValGlyValGlyIleTyr
1970 1980

5950 5955

CTGCTCCCCAACCGA (SEQ. ID. NO. 5)
-----+-----
LeuLeuProAsnArg (SEQ. ID. NO. 6)
1985

FIG. 50

1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
 51 GAGACGGTCA CAGCTTGCT GTAAAGCGGAT GCCGGGAGCA GACAAGCCCG
 101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
 151 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
 201 CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCA
 251 TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
 301 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
 351 AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCCGCGTT
 401 ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG
 451 CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
 501 CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
 551 GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 601 TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
 651 ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
 701 GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC
 751 ACGGGGATT TCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTGTTT
 801 GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA
 851 TTGACGAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
 901 AGCTCGTTA GTGAACCGTC AGATGCCCTG GAGACGCCAT CCACGCTGTT
 951 TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA
 1001 CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC
 1051 CTATAGACTC TATAGGCACA CCCCTTGCG TCTTATGCAT GCTATACTGT
 1101 TTTGGCTTG GGGCTATAC ACCCCCCGTT CCTTATGCTA TAGGTGATGG
 1151 TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
 1201 TATTGGTGAC GATACTTCC ATTACTAATC CATAACATGG CTCTTGCCA
 1251 CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGCTTTC AGAGACTGAC
 1301 ACGGACTCTG TATTTTTACA GGATGGGTC CCATTTATTA TTTACAAATT
 1351 CACATATACA ACAACGCCGT CCCCCGTGCC CGCAGTTTT ATTAAACATA
 1401 GCGTGGGATC TCCACCGAA TCTCGGGTAC GTGTTCCCGA CATGGGCTCT
 1451 TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCTCC
 1501 AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCCTAACAG TGGAGGCCAG
 1551 ACTTAGGCAC AGCACAATGC CCACCAACAC CAGTGTCCCG CACAAGGCCG
 1601 TGGCGGTAGG GTATGTGCT GAAAATGAGC GTGGAGATTG GGCTCGCACG
 1651 GCTGACGCAG ATGGAAGACT TAAGGCAGCG GCAGAAGAAG ATGCAGGCAG
 1701 CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC GTTGGCGGTGC
 1751 TGTAAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
 1801 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT
 1851 GGGTCTTTTC TGCAGTCACC GTCCTTAGAT CTAGGTACCA GATATCAGAA
 1901 TTCAGTCGAC AGCGGCCGCG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC
 1951 TGTGTTTGC CCCTCCCCCG TGCCCTCCTT GACCCTGGAA GGTGCCACTC
 2001 CCACGTGCTT TTCTAAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT
 2051 AGGTGTCAATT CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGAA

40/92

2101 GGATTGGAA GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG
2151 CCGCTGCGGC CAGGTGCTGA AGAATTGACC CGGTTCCCTCC TGGGCCAGAA
2201 AGAACAGGC ACATCCCTT CTCTGTGACA CACCCCTGTCC ACGCCCTGG
2251 TTCTTAGTTC CAGCCCCACT CATAGGACAC TCATAGCTCA GGAGGGCTCC
2301 GCCTCAATC CCACCCGCTA AAGTACTTGG AGCGGTCTCT CCCTCCCTCA
2351 TCAGCCCACC AAACCAAACCC TAGCCTCCAA GAGTGGGAAG AAATAAAGC
2401 AAGATAGGCT ATTAAGTGCA GAGGGAGAGA AAATGCCTCC AACATGTGAG
2451 GAAGTAATGA GAGAAATCAT AGAATTCTT CCGCTTCCCTC GCTCACTGAC
2501 TCGCTGCGCT CGGTGCGTTG GCTGCGGGA GCGGTATCAG CTCACTCAA
2551 GGGGTAATA CGGTTATCCA CAGAACTCAGG GGATAACGCA GGAAAGAAC
2601 TGTGAGCAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG
2651 CTGGCGTTT TCCATAGGCT CCGCCCCCT GACGAGCATIC ACAAAAATCG
2701 ACGCTCAAGT CAGAGGTGGC GAAACCCGAC AGGACTATAA AGATACCAAGG
2751 CGTTTCCCCC TGGAGCTCC CTCGTGCGCT CTCCTGTTCC GACCCTGCC
2801 CTTACCGGAT ACCTGTCCGC CTTTCTCCCT TCGGGAAAGCG TGGCGCTTC
2851 TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA
2901 AGCTGGCTG TGTGCACGAA CCCCCCGTTC AGCCCGACCG CTGCGCCTTA
2951 TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC
3001 ACTGGCAGCA GCCACTGGTA ACAGGATTAG CAGAGCGAGG TATGTAGGCG
3051 GTGCTACAGA GTTCTTGAAG TGGTGGCCTA ACTACGGCTA CACTAGAAGA
3101 ACAGTATTTG GTATCTGCC TCTGCTGAAG CCAGTTACCT TCGGAAAAAG
3151 AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT AGCGGTGGTT
3201 TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG ATCTCAAGAA
3251 GATCCTTGA TCTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAAC
3301 ACGTTAAGGG ATTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA
3351 TCCTTTAAA TTAAAAATGA AGTTTAAAT CAATCTAAAG TATATATGAG
3401 TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG CACCTATCTC
3451 AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCCTGACTC GGGGGGGGG
3501 GGCCTGAGG TCTGCCCTGT GAAGAAGGTG TTGCTGACTC ATACCAGGCC
3551 TGAATCGCCC CATCATCCAG CCAGAAAGTG AGGGAGCCAC GGTTGATGAG
3601 AGCTTGTG TAGGTGGACC AGTTGGTGTAT TTGAACTTT TGCTTGCCA
3651 CGGAACGGTC TCGTTGTCG GGAAGATGCG TGATCTGATC CTTCAACTCA
3701 GCAAAAGTTC GATTTATTCA ACAAAAGCCGC CGTCCCCGTCA AGTCAGCGTA
3751 ATGCTCTGCC AGTGTACAA CCAATTAACC AATTCTGATT AGAAAAAACTC
3801 ATCGAGCATIC AAATGAAACT GCAATTATT CATATCAGGA TTATCAATAC
3851 CATATTTTG AAAAAGCCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG
3901 CAGTTCCATA GGATGGCAAG ATCCTGGTAT CGGTCTGCCA TTCCGACTCG
3951 TCCAACATCA ATACAACCTA TTAATTCCC CTCGTAAAA ATAAGGTTAT
4001 CAAGTGAGAA ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAA
4051 AGCTTATGCA TTTCTTCCA GACTTGTCA ACAGGCCAGC CATTACGCTC
4101 GTCATCAAAA TCACTCGCAT CAACCAAACCC GTTATTCAATT CGTGATTGCG
4151 CCTGAGCGAG AGGAATACG CGATCGCTGT TAAAAGGACA ATTACAAACA

FIG. 6B

41/92

4201 GGAATCGAAT GCAACC GGCG CAGGAACACT GCCAGCGCAT CAACAATATT
4251 TTCACCTGAA TCAGGATATT CTTCTAATAC CTGGAATGCT GTTTTCCCGG
4301 GGATCGCAGT GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC
4351 TTGATGGTCG GAAGAGGCAT AAATTCCGTC AGCCAGTTA GTCTGACCAT
4401 CTCATCTGTA ACATCATTGG CAACGCTACC TTTGCCATGT TTCAGAAAACA
4451 ACTCTGGCGC ATCGGGCTTC CCATACAATC GATAGATTGT CGCACCTGAT
4501 TGCCCGACAT TATCGCGAGC CCATTTATAC CCATATAAAAT CAGCATCCAT
4551 GTTCCAATT AATCGCGGC TCGAGCAAGA CGTTTCCCGT TGAATATGGC
4601 TCATAACACC CCTTGTTATTCTGTTTATGT AAGCAGACAG TTTTATTGTT
4651 CATGATGATA TATTTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA
4701 CACAACGTGG CTTTCCCCCC CCCCCCATTAA TTGAAGCATT TATCAGGGTT
4751 ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA
4801 ATAGGGGTTTC CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA
4851 AACCAATTATT ATCATGACAT TAACCTATAA AAATAGGCGT ATCACGAGGC
4901 CCTTTCGTC

FIG. 6C

42/92

1 CATCATCAAT AATATAACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
 61 TTGTGACGTG GCGCGGGCG TGGGAACGGG GCGGGTGACG TAGTAGTGTG GCGGAAGTGT
 121 GATGTTGTAA GTGTGGCGGA ACACATGTAA GCGCCGGATG TGGTAAAAGT GACGTTTTTG
 181 GTGTGCGCCG GTGTACACGG GAAGTGCACAA TTTTCGCGCG GTTTTAGGCG GATGTTGTAG
 241 TAAATTTGGG CGTAACCAAG TAATATTGG CCATTTCGC GGGAAAATG AATAAGAGGA
 301 AGTGAATCT GAATAATCT GTGTTACTCA TAGCGCGTAA TATTTGTCTA GGGCCGCGGG
 361 GACTTGACC GTTTACGTGG AGACTCGCCC AGGTGTTTTT CTCAGGTGTT TTCCCGCGTTC
 421 CGGGTCAAAG TTGGCGTTT ATTATTATAG TCAGCTGACG CGCAGTGTAT TTATACCCGG
 481 TGAGTTCCCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTCTCC TCCGAGCCGC
 541 TCCGACACCG GGACTGAAA TGAGACATAT TATCTGCCAC GGAGGTGTTA TTACCGAAGA
 601 AATGGCCGCC AGTCTTTGG ACCAGCTGAT CGAAGAGGTA CTGGCTGATA ATCTTCCACC
 661 TCCTAGCCAT TTTGAACCAC CTACCCCTCA CGAACTGTAT GATTTAGACG TGACGGCCCC
 721 CGAAGATCCC AACGAGGAGG CGGTTTCGCA GATTTTCCC GAGTCGTAA TGTTGGCGGT
 781 GCAGGAAGGG ATTGACTTAT TCACTTTCC GCGGGCGCC GGTCTCCGG AGCCGCCTCA
 841 CCTTTCCCGG CAGCCCGAGC AGCCGGAGCA GAGAGCCTTG GGTCCGGTTT CTATGCCAAA
 901 CCTTGTGCCG GAGGTGATCG ATCTTACCTG CCACGAGGCT GGCTTCCAC CCAGTGACGA
 961 CGAGGATGAA GAGGGTGAGG AGTTTGTGTT AGATTATGT GAGCACCCCG GGCACGGTTG
 1021 CAGGTCTTGT CATTATCACCC GGAGGAATAC GGGGGACCCA GATATTATGT GTTCGCTTTG
 1081 CTATATGAGG ACCTGTGGCA TGTTTGTCTA CAGTAAGTGA AAAATTATGG GCAGTGGGTG
 1141 ATAGAGTGGT GGGTTTGGTG TGTTAATT TTTTTAATT TTTACAGTTT TGTTGGTTAA
 1201 AGAATTTTGT ATTGTGATTT TTTAAAAGGT CCTGTGTCTG AACCTGAGCC TGAGCCCGAG
 1261 CCAGAACCGG AGCCTGCAAG ACCTACCCGG CGTCCTAAAT TGGTGCCTGC TATCCTGAGA
 1321 CGCCCGACAT CACCTGTGTC TAGAGAATGC AATAGTAGTA CGGATAGCTG TGACTCCGGT
 1381 CCTCTAAACA CACCTCCTGA GATACACCCG GTGGTCCCGC TGTGCCCAT TAAACCAGTT
 1441 GCCGTGAGAG TTGGTGGCG TCGCCAGGCT GTGGAATGTA TCGAGGACTT GCTTAACGAG
 1501 TCTGGGCAAC CTTTGGACTT GAGCTGTAAA CGCCCCAGGC CATAAGGTGT AAACCTGTGA
 1561 TTGCGTGTGT GGTTAACGCC TTTGTTGCT GAATGAGTTG ATGTAAGTTT AATAAAGGGT
 1621 GAGATAATGT TTAACTTGCA TGGCGTGTAA AATGGGGCGG GGCTTAAAGG GTATATAATG
 1681 CGCCGTGGGC TAATCTTGGT TACATCTGAC CTCATGGAGG CTTGGGAGTG TTGGGAAGAT
 1741 TTTCTGTCTG TGCGTAACCT GCTGGAACAG AGCTCTAACAA GTACCTCTTG GTTTGGAGG
 1801 TTTCTGTGGG GCTCCTCCCA GGAAAGTTA GTCTGCAGAA TTAAGGAGGA TTACAAGTGG
 1861 GAATTGAAAG AGCTTTGAA ATCCTGTGGT GAGCTGTTG ATTCTTGAA TCTGGGTAC
 1921 CAGGCCTTT TCCAAGAGAA GGTCATCAAG ACTTTGGATT TTTCCACACC GGGGCGCGCT
 1981 GCGGCTGCTG TTGCTTTT GAGTTTATA AAGGATAAAAT GGAGCGAAGA AACCCATCTG
 2041 AGCGGGGGGT ACCTGCTGGA TTTCTGGCC ATGCATCTGT GGAGAGCGGT GGTGAGACAC
 2101 AAGAACGCC TGCTACTGTGTT GTCTTCCGTC CGCCCCGGCAA TAATACCGAC GGAGGAGCAA
 2161 CAGCAGGAGG AAGCCAGGCG CGGGCGCGG CAGGAGCAGA GCCCCATGGAA CCCGAGAGCC
 2221 GGCCCTGGACC CTCGGGAATG AATGTTGTAC AGGTGGCTGA ACTGTTCCA GAACTGAGAC
 2281 GCATTTAAC CATTAAACGAG GATGGGCAGG GGCTAAAGGG GTAAAGAAG GAGCGGGGGG
 2341 CTTCTGAGGC TACAGAGGAG GCTAGGAATC TAACTTTAG CTTAATGACC AGACACCGTC
 2401 CTGAGTGTGT TACTTTTCAG CAGATTAAGG ATAATTGCGC TAATGAGCTT GATCTGCTGG
 2461 CGCAGAAAGTA TTCCATAGAG CAGCTGACCA CTTACTGGCT GCAGGCCAGGG GATGATTGG

FIG. 7A

43/92

2521 AGGAGGCTAT TAGGGTATAT GCAAAGGTGG CACTTAGGCC AGATTGCAAG TACAAGATTA
 2581 GCAAACCTGT AAATATCAGG AATTGTTGCT ACATTTCTGG GAACGGGCC GAGGTGGAGA
 2641 TAGATACGGÀ GGATAGGGTG GCCTTTAGAT GTAGCATGAT AAATATGTGG CCGGGGGTGC
 2701 TTGGCÀTGGÀ CGGGGTGGTT ATTATGAATG TGAGGTTAC TGGTCCCAAT TTAGCGGTA
 2761 CGGTTTCCTT GGCAATACC AATCTTATCC TACACGGTGT AAGCTCTAT GGGTTTAACA
 2821 ATACCTGTGT GGAAGCCTGG ACCGATGTAA GGGTTCGGGG CTGTCCCTTT TACTGCTGCT
 2881 GGAAGGGGGT GGTGTGTCGC CCCAAAAGCA GGGCTTCAT TAAGAAATGC CTGTTTGAAA
 2941 GGTGTACCTT GGGTATCCTG TCTGAGGTAA ACTCCAGGGT GCGCCACAAT GTGGCCTCCG
 3001 ACTGTGGTTG CTTTATGCTA GTGAAAGCG TGGCTGTGAT TAAGCATAAC ATGGTGTGTC
 3061 GCAACTGCGA GGACAGGGCC TCTCAGATGC TGACCTGCTC GGACGGCAAC TGTCACTTGC
 3121 TGAAGACCAT TCACGTAGCC AGCCACTCTC GCAAGGCCTG GCCAGTGTGTT GAGCACAA
 3181 TACTGACCCG CTGTTCCCTG CATTGGTAA ACAGGAGGGG GGTGTTCCCTA CCTTACCAAT
 3241 GCAATTGAG TCACACTAAG ATATTGCTTG AGCCCGAGAG CATGTCCAAG GTGAACCTGA
 3301 ACGGGGTGTGTT TGACATGACC ATGAAGATCT GGAAGGTGCT GAGGTACGAT GAGACCCGCA
 3361 CCAGGTGCAG ACCCTGCGAG TGTGGCGTA AACATATTAG GAAACCAGCCT GTGATGCTGG
 3421 ATGTGACCGA GGAGCTGAGG CCCGATCACT TGGTGCTGGC CTGCACCCCGC GCTGAGTTG
 3481 GCTCTAGCGA TGAAGATACA GATTGAGGTA CTGAAATGTG TGGGCGTGGC TTAAGGGTGG
 3541 GAAAGAATAT ATAAGGTGGG GGTCTCATGT AGTTTTGTAT CTGTTTGCA GCAGCCGCCG
 3601 CCATGAGCGC CAACTCGTT GATGGAAGCA TTGTGAGCTC ATATTGACA ACGCGCATGC
 3661 CCCCATGGGC CGGGGTGCGT CAGAATGTGA TGGGCTCCAG CATTGATGGT CGCCCCGTCC
 3721 TGCCCGAAA CTCTACTACC TTGACCTACG AGACCCTGTC TGGAACGCCG TTGGAGACTG
 3781 CAGCCCTCCGC CGCCGCTTCA GCGCGTGCAG CCACCGCCCCG CGGGATTGTG ACTGACTTTG
 3841 CTTTCCTGAG CCCGCTTGCA AGCAGTGCAG CTTCCCGTTC ATCCGCCCGC GATGACAAGT
 3901 TGACGGCTCT TTTGGCACAA TTGGATTCTT TGACCCGGGA ACTTAATGTC GTTTCTCAGC
 3961 AGCTGTTGGA TCTGCGCCAG CAGGTTCTG CCCTGAAGGC TTCCCTCCCT CCCAATGCCG
 4021 TTTAAAACAT AAATAAAAAC CAGACTCTGT TTGGATTGAG ATCAAGCAAG TGTCTTGCTG
 4081 TCTTATTTA GGGGTTTGC CGCCGCGGTAA GGCCCGGGAC CAGCGGTCTC GGTGTTGAG
 4141 GGTCTGTGT ATTTTTCCA GGACGTGGTA AAGGTGACTC TGGATGTTCA GATACATGGG
 4201 CATAAGCCCG TCTCTGGGGT GGAGGTAGCA CCACTGCAGA GCTTCATGCT CGGGGGTGGT
 4261 GTTGTAGATG ATCCAGTCGT AGCAGGGAGCG CTGGGCGTGG TGCCTAAAAA TGTCTTCA
 4321 TAGCAAGCTG ATTGCCAGGG GCAGGCCCTT GGTGTAAGTG TTTACAAAGC GGTAAAGCTG
 4381 GGATGGGTGC ATACGTGGGG ATATGAGATG CATCTGGAC TGTATTTTA GTTGGCTAT
 4441 GTTCCCAGCC ATATCCCTCC GGGGATTCAAT GTTGTGCAGA ACCACCAAGCA CAGTGTATCC
 4501 GGTGCACTTG GGAAATTGT CATGTAGCTT AGAAGGAAAT GCGTGGAAAGA ACTTGGAGAC
 4561 GCCCTTGTGA CCTCCAAGAT TTTCCATGCA TTCGTCCATA ATGATGGCAA TGGGCCACG
 4621 GGCAGCGGCC TGGGCGAAGA TATTCCTGGG ATCACTAACG TCATAGTTGT GTTCCAGGAT
 4681 GAGATCGTCA TAGGCCATT TTACAAAGCG CGGGCGGAGG GTGCCAGACT CGGGTATAAT
 4741 GGTTCCATCC GGCCCAGGGG CGTAGTGTACCT CTCACAGATT TGCATTCCC ACGCTTTGAG
 4801 TTCAGATGGG GGGATCATGT CTACCTGCAGG GGCAGTGAAG AAAACCGTTT CGGGGGTAGG
 4861 GGAGATCAGC TGGGAAGAAA GCAGGTTCCCT AAGCAGCTGC GACTTACCGC AGCCGGTGGG
 4921 CCCGTAATAC ACACCTATTA CCGGCTGCAA CTGGTAGTTA AGAGAGCTGC AGCTGCCGTC
 4981 ATCCCTGAGC AGGGGGGCCA CTTCGTTAAG CATGTCCCTG ACTTGCATGT TTTCCCTGAC

FIG. 7B

44/92

5041 CAAATCCGCC AGAAGGCAGT CGCCGCCAG CGATAGCAGT TCTTGCAAGG AAGCAAAGTT
 5101 TTTCAACGGT TTGAGGCCGT CGCCGTAGG CATGTTTTG AGCGTTGAC CAAGCAGTTC
 5161 CAGGCGGTCC CACAGCTCGG TCACGTGCTC TACGGCATCT CGATCCAGCA TATCTCCTCG
 5221 TTTCGCGGGT TGGGGCGGCT TTCGCTGTAC GGCACTAGTC GGTGCTCGTC CAGACGGGCC
 5281 AGGGTCATGT CTTTCCACGG GCGCAGGGTC CTCGTCAGCG TAGTCTGGGT CACGGTGAAG
 5341 GGGTGCCTCG CGGGTTGCAG GCTGGCCAGG GTGCCTTGA GGCTGGTCCT GCTGGTGCCTG
 5401 AAGCGCTGCC GGTCTTCGCC CTGCGCGTCG GCCAGGTAGC ATTTGACCAT GGTGTATAG
 5461 TCCAGCCCCCT CCGCGCGCTG GCCCTTGGCG CGCAGCTTGC CCTTGGAGGA GGCAGCCGCAC
 5521 GAGGGGCAGT GCAGACTTT AAGGGCGTAG AGCTTGGCG CGAGAAATAC CGATTCCGGG
 5581 GAGTAGGCAT CCGCGCCGCA GGCCCCCGAG ACGGTCTCGC ATTCCACGAG CCAGGTGAGC
 5641 TCTGGCCGTT CGGGGTAAA AACCAGGTTT CCCCCATGCT TTTTGATGCG TTTCTTACCT
 5701 CTGGTTTCCA TGAGCCGGTG TCCACGCTCG GTGACGAAA GGCTGTCCTGT GTCCCCGTAT
 5761 ACAGACTTGA GAGGCCTGTC CTCGAGCGGT GTTCCGCGGT CCTCCTCGTA TAGAAACTCG
 5821 GACCACTCTG AGACGAAGGC TCGCGTCCAG GCCAGCACGA AGGAGGCTAA GTGGGAGGGGG
 5881 TAGCGGTCGT TGTCCACTAG GGGTCCACT CGCTCCAGGG TGTGAAGACA CATGTCGCC
 5941 TCTCGGCAT CAAGGAAGGT GATTGGTTA TAGGTGTAGG CCACGTGACC GGGTGTTCCT
 6001 GAAGGGGGGC TATAAAAGGG GGTGGGGCG CGTTCGTCCCT CACTCTCTTC CGCATCGCTG
 6061 TCTGCGAGGG CCAGCTGTTG GGGTGAGTAC TCCCTCTCAA AAGCGGGCAT GACTTCTGCG
 6121 CTAAGATTGT CAGTTTCAA AAACGAGGAG GATTGATAT TCACCTGGCC CGCGGTGATG
 6181 CCTTGAGGG TGGCCCGTC CATCTGGTCA GAAAAGACAA TCTTTTGTT GTCAAGCTTG
 6241 GTGGCAAACG ACCCGTAGAG GGCGTTGGAC AGCAACTTGG CGATGGAGCG CAGGGTTTGG
 6301 TTTTGTCGC GATCGGCCG CTCCCTGGCC GCGATGTTA GCTGCACTGTA TTCGCGCGCA
 6361 ACGCACCGCC ATTCCGGAAA GACGGTGGTG CGCTCGTCGG GCACTAGGTG CACCGGCCAA
 6421 CGCGGTTGT GCAGGGTGAC AAGGTCAACG CTGGTGGCTA CCTCTCCGCG TAGGGCCTCG
 6481 TTGGTCCAGC AGAGGCGGCC GCCCTTGCAG GAGCAGAATG GCGGTAGTGG GTCTAGCTGC
 6541 GTCTCGTCGG GGGGGTCTGC GTCCACGGTA AAGACCCCCGG GCAGCAGGCG CGCGTCGAAG
 6601 TAGTCTATCT TGCATCCTTG CAAGTCTAGG GCCTGCTGCC ATGCGGGCG GGCAAGCGCG
 6661 CGCTCGTATG GGTTGAGTGG GGGACCCCAT GGCAATGGGGT GGGTGAGCGC GGAGGGCGTAC
 6721 ATGCCGAAA TGTCGAAAC GTAGAGGGC TCTCTGAGTA TTCCAAGATA TGTAGGGTAG
 6781 CATCTTCCAC CGCGGATGCT GGCGCGCAG TAATCGTATA GTTCGTCGCA GGGAGCGAGG
 6841 AGGTGGGAC CGAGGTTGCT ACGGGCGGGC TGCTCGTC GGAAGACTAT CTGCCTGAAG
 6901 ATGGCATGTG AGTTGGATGA TATGGTTGGA CGCTGGAAGA CGTTGAAGCT GGCGTCTGTG
 6961 AGACCTACCG CGTCACGCAC GAAGGAGGCG TAGGAGTCGC GCAGCTTGT GACCAGCTCG
 7021 GCGGTGACCT GCACGTCTAG GGCGCAGTAG TCCAGGGTTT CCTTGATGAT GTCATACTTA
 7081 TCCTGTCCTT TTTTTTCCA CAGCTCGCGG TTGAGGACAA ACTCTCGCG GTCTTCCAG
 7141 TACTCTTGGGA TCGGAAACCC GTCGGCCCTCC GAACGGTAAG AGCCTAGCAT GTAGAACTGG
 7201 TTGACGGCCT GGTAGGCAGA GCATCCCTTT TCTACGGGTA CGCGTATGC CTGCGCGGCC
 7261 TTCCGGAGCG AGGTGTGGGT GAGCGCAAAG GTGTCCCTAA CCATGACTTT GAGGTACTGG
 7321 TATTTGAAGT CAGTGTGTC GCATCCGCC TGCTCCCAGA GCAAAAGTC CGTGGCTTT
 7381 TTGGAACGCG GGTTTGGCAG GGCGAAGGTG ACATCGTGA AGAGTATCTT TCCCGCGCGA
 7441 GGCATAAAAGT TGCCTGTGAT GCGGAAGGGT CCCGGCACCT CGGAACGGTT GTTAATTACC
 7501 TGGCGGGCGA GCACGATCTC GTCAAAGCCG TTGATGTGT GGCCCACAAT GTAAAGTCCC

FIG. 7C

45/92

7561 AAGAAGCGCG GGATGCCCTT GATGGAAGGC AATTTTTAA GTTCCCTCGTA GGTGAGCTCT
 7621 TCAGGGGAGC TGAGCCCCTG CTCTGAAAGG GCCCAGTCG CAAGATGAGG GTTGGAAAGCG
 7681 ACGAATGAGC TCCACAGGTC ACGGCCATT AGCATTGCA GGTGGTCGCG AAAGGTCTTA
 7741 AACTGGCGAC CTATGGCCAT TTTTCTGGG GTGATGCAGT AGAAGGTAAG CGGGTCTTGT
 7801 TCCCAGCGGT CCCATCCAAG GTCCCGGGCT AGGTCTCGCG CGGCCGTCAC TAGAGGCTCA
 7861 TCTCCGCCGA ACTTCATGAC CAGCATGAAG GGCACGAGCT GCTTCCAAA GGCCCCCATC
 7921 CAAGTATAGG TCTCTACATC GTAGGTGACA AAGAGACGCT CGGTGCGAGG ATGCGAGCCG
 7981 ATCAGGAAGA ACTGGATCTC CCGCCACCAG TTGGAGGAGT GGCTGTTGAT GTGGTGAAAG
 8041 TAGAAGTCCC TGCGACGGC CGAACACTCG TGCTGGCTT TGAAAAAACG TGCGCAGTAC
 8101 TGCGAGCGGT GCACGGGCTG TACATCCTGC ACGAGGTTGA CCTGACGACC GCGCACAAAGG
 8161 AAGCAGAGTG GGAATTGAG CCCCTCGCCT GGCAGGGTTG GCTGGTGGTC TTCTACTTCG
 8221 GCTGCTTGTC CTTGACCGTC TGGCTGCTCG AGGGGAGTTA CGGTGGATCG GACCACCACG
 8281 CCGCGCGAGC CCAAAGTCCA GATGTCGCG CGCGCGGGTC GGAGCTTGAT GACAACATCG
 8341 CGCAGATGGG AGCTGTCCAT GGTCTGGAGC TCCCGCGGGC TCAGGTCAGG CGGGAGCTCC
 8401 TGCAGGTTTA CCTCGCATAG CCGGGTCAGG GCGCGGGCTA GGTCCAGGTG ATACCTGATT
 8461 TCCAGGGGCT GGTTGGTGGC GGCCTGATG GCTTGCAAGA GGCGCATCC CGCGGGCGCG
 8521 ACTACGGTAC CGCGCGGGCG GCGGTGGGCC GCGGGGGTGT CCTTGGATGA TGCATCTAAA
 8581 AGCGGTGACG CGGGCGGGCC CCCGGAGGTA GGGGGGGCTC GGGACCCGCC GGGAGAGGGG
 8641 GCAGGGGCAC GTCGCGCGCG CGCGCGGGCA GGAGCTGGT GTCGCGCGGG AGGTTGCTGG
 8701 CGAACCGGAC GACCGCGGG TTGATCTCCT GAATCTGGCG CCTCTGCGTG AAGACGACGG
 8761 GCCCGGTGAG CTTGAACCTG AAAGAGAGTT CGACAGAACATC AATTTCGGTG TCGTTGACGG
 8821 CGGCCCTGGCG CAAAATCTCC TGACAGTCTC CTGAGTTGTC TTGATAGGCG ATCTCGGCCA
 8881 TGAACGTCTC GATCTCTTC TCCTGGAGAT CTCCCGCTCC GGCTCGCTCC ACGGTGGCGG
 8941 CGAGGTCGTT GGAGATGCGG GCCATGAGCT GCGAGAAGGC GTTGAGGCCT CCCTCGTTCC
 9001 AGACCGGGCT GTAGACCAAG CCCCCCTCGG CATCGCGGGC GCGCATGACC ACCTGCGCGA
 9061 GATTGAGCTC CACGTGCCGG GCGAAGACGG CGTAGTTTCG CAGGCCCTGA AAGAGGTAGT
 9121 TGAGGGTGGT GGCGGTGTGT TCTGCCACGA AGAAGTACAT AACCCAGCGC CGCAACGTGG
 9181 ATTGTTGAT ATCCCCCAAG GCCTCAAGGC GCTCCATGGC CTCGTAGAAG TCCACGGCGA
 9241 AGTTGAAAAA CTGGGAGTTG CGCGCGGACA CGGTTAACCTC CTCCCTCCAGA AGACGGATGA
 9301 GCTCGGGCAGC AGTGTGCGCC ACCTCGCGCT CAAAGGCTAC AGGGGCCCT TCTTCTTCTT
 9361 CAATCTCTC TTCCATAAGG GCCTCCCCCTT CTTCTTCTTC TGGCGGGCGGT GGGGGAGGGG
 9421 GGACACGGCG GCGACGACGG CGCACCGGGG GCGGGTCGAC AAAGCGCTCG ATCATCTCC
 9481 CGCGGGCAGC GCGCATGGTC TCGGTGACGG CGCGGGCGTT CTCGCGGGGG CGCAGTTGGA
 9541 AGACGCCGCC CGTCATGTCC CGGTTATGGG TTGGCGGGGG GCTGCCGTGC GGCAGGGATA
 9601 CGGGCCTAAC GATGCATCTC AACAAATTGTT GTGTAGGTAC TCCGCCACCG AGGGACCTGA
 9661 GCGAGTCCGC ATCGACCGGA TCGGAAAACC TCTCGAGAAA GGCCTCAAC CAGTCACAGT
 9721 CGCAAGGTAG GCTGAGCACC GTGGCGGGCG GCAGCGGGCG GCGGTGGGG TTGTTCTGG
 9781 CGGAGGTGCT GCTGATGATG TAATTAAAGT AGGCGGTCTT GAGACGGCGG ATGGTCGACA
 9841 GAAGCACCACAT GTCCTTGGGT CGGGCCTGCT GAATGCGCAG GCGGTGGGCC ATGCCCCAGG
 9901 CTTCGTTTGT ACATCGGCCG AGGTCTTGT AGTAGTCTTG CATGAGCCTT TCTACCGGCA
 9961 CTTCTTCTTC TCCTTCTCT TGTCCCTGCAT CTCTTGCATC TATCGCTGCG CGGGCGGGCG
 10021 AGTTGGCG TAGGTGGCGC CCTCTTCTC CCATGCGTGT GACCCGAAG CCCCTCATCG

FIG. 7D

46/92

10081 GCTGAAGCAG GGCCAGGTCG GCGACAACGC GCTCGGCTAA TATGGCCTGC TGCACCTGCG
 10141 TGAGGGTAGA CTGGAAGTCG TCCATGTCCA CAAAGCGGTG GTATGCGCCC GTGTTGATGG
 10201 TGTAAGTGC A GTTGGCCATA ACGGACCACT TAACCGTCTG GTGACCCGGC TGGAGAGCT
 10261 CGGTGTACCT GAGACGCGAG TAAGCCCTTG AGTCAAAGAC GTAGTCGTTG CAAGTCCGCA
 10321 CCAGGTACTG GTATCCCACC AAAAAGTGC GCGGCGGCTG GCGGTAGAGG GGCCAGCGTA
 10381 GGGTGGCCGG GGCTCCGGG GCGAGGTCTT CCAACATAAG GCGATGATAT CCGTAGATGT
 10441 ACCTGGACAT CCAGGTGATG CCGGCGGCG TGTTGGAGGC GCGCGAAAG TCACGGACGC
 10501 GGTTCCAGAT GTTGCAGC GGCAAAAGT GCTCCATGGT CGGGACGCTC TGGCCGGTCA
 10561 GGCGCGCCGA GTCTGGTACG CTCTAGACCG TGCAAAAGGA GAGCCTGTAAC GCGGGCACTC
 10621 TTCCGTGGTC TGTTGGATAA ATTGCAAGG GTATCATGGC GGACGACCGG GGTCGAACC
 10681 CCGGATCCGG CCGTCCGCCG TGATCCATGC GGTTACCGCC CGCGTGTGCA ACCCAGGTGT
 10741 GCGACGTCAG ACAACGGGG AGCGCTCCTT TTGGCTTCCT TCCAGGCCG GCGGATGCTG
 10801 CGCTAGCTTT TTGCGCAACT GGCGCGCGC GCGCTAAGCG GTTAGGCTGG AAAGCGAAAG
 10861 CATTAAAGTGG CTCGCTCCCT GTAGCCGGAG GGTTATTTTC CAAGGGTTGA GTCGCGGGAC
 10921 CCCCGGTTCG AGTCTCGGC CGGCCGACT GCGCGAACG GGGGTTGCC TCCCCGTCAT
 10981 GCAAGACCCC GCTTGCAAAT TCCTCCGGAA ACAGGGACGA GCCCCCTTTTT TGCTTMTCCC
 11041 AGATGCATCC GGTGCTGCGG CAGATGCGCC CCCCTCCTCA GCAGCGGCAA GAGCAAGAGC
 11101 AGCGGCAGAC ATGCAGGGCA CCCTCCCCCTT CTCCCTACCGC GTCAGGAGGG GCAACATCCG
 11161 CGGCTGACGC CGCGCAGAT GGTGATTACG AACCCCGCG GCGCCGGACC CGGCACTACT
 11221 TGGACTTGGG GGAGGGCGAG GGCTGGCGC GGCTAGGAGC GCCCCTCCT GAGCGACACC
 11281 CAAGGGTGCA GCTGAAGCGT GACACGCGC AGGCGTACGT GCGCGGGCAG AACCTGTTTC
 11341 GCGACCGCGA GGGAGAGGAG CCCGAGGAGA TGCGGGATCG AAAGTTCCAT GCAGGGCGCG
 11401 AGTTGCGGCA TGGCCTGAAC CGCGAGCGGT TGCTGCGCA GGAGGACTTT GAGCCCGACG
 11461 CGCGGACCGG GATTAGTCCC GCGCGCGAC ACGTGGCGGC CGCCGACCTG GTAACCGCGT
 11521 ACGAGCAGAC GGTGAACCAAG GAGATTAAC TTCAAAAAAG CTTTAACAAAC CACGTGCGCA
 11581 CGCTTGTGGC GCGCGAGGAG GTGCGTATAG GACTGATGCA TCTGTGGGAC TTGTAAGCG
 11641 CGCTGGAGCA AAACCCAAAT AGCAAGCCGC TCATGGCGA GCTGTTCCCTT ATAGTGCAGC
 11701 ACAGCAGGGCA CAACGAGGCA TTCAGGGATG CGCTGCTAAA CATAGTAGAG CCCGAGGGCC
 11761 GCTGGCTGCT CGATTGATA AACATTCTGC AGAGCATAGT GGTGCAGGAG CGCAGCTTGA
 11821 GCCTGGCTGA CAAGGGGCC GCCATTAAC ATTCCATGCT CAGTCTGGC AAGTTTTACG
 11881 CCCGCAAGAT ATACCATACC CCTTACGTC CCATAGACAA GGAGGTAAG ATCGAGGGGT
 11941 TCTACATGCG CATGGCGCTG AAGGTGCTTA CCTTGAGCGA CGACCTGGGC GTTTATCGCA
 12001 ACGAGCGCAT CCACAAGGCC GTGAGCGTGA GCGGGCGCG CGAGCTCAGC GACCGCGAGC
 12061 TGATGCACAG CCTGCAAAGG GCGCTGGCTG GCACGGGAG CGGGGATAGA GAGGCCGAGT
 12121 CCTACTTTGA CGCGGGCGCT GACCTGCGCT GGGCCCCAAG CCGACGCGCC CTGGAGGCAG
 12181 CTGGGGCCGG ACCTGGGCTG GCGGTGGCAC CGCGGGCGC TGCAACGTC GCGGGCGTGG
 12241 AGGAATATGA CGAGGACGAT GAGTACGAGC CAGAGGACGG CGAGTACTAA GCGGTGATGT
 12301 TTCTGATCAG ATGATGCAAG ACGCAACGGA CCCGGCGGTG CGGGCGGC TGCAAGAGCCA
 12361 GCGTCCGGC CTTAACCTCA CGGACGACTG GCGCCAGGTC ATGGACCGCA TCATGTCGCT
 12421 GACTGCGCG AACCCCTGACG CGTTCCGGCA GCAGCCGAG GCCAACCGGC TCTCCGCAAT
 12481 TCTGGAAGCG GTGGTCCGG CGCGCGAAA CCCCCACGAC GAGAAGGTGC TGGCGATCGT
 12541 AAACGCGCTG GCGAAAACA GGGCCATCCG GCGGATGAG GCGGGCCTGG TCTACGACGC

FIG. 7E

47/92

12601 GCTGCTTCAG CGCGTGGCTC GTTACAACAG CAGCAACGTG CAGACCAACC TGGACCGGGT
 12661 GGTGGGGGAT GTGCGCGAGG CCGTGGCGA GCGTGAGCGC GCGCAGCAGC AGGGCAACCT
 12721 GGGCTCCATG GTTGCACCAAACG CCGCTTCCT GAGTACACAG CCCGCCAACG TGCCGCGGG
 12781 ACAGGAGGAC TACACCAACT TTGTGAGCGC ACTGCGGCTA ATGGTGAATG AGACACCGCA
 12841 AAGTGAGGTG TATCAGTCCG GGCCAGACTA TTTTTCCAG ACCAGTAGAC AAGGCCTGCA
 12901 GACCGTAAAC CTGAGCCAGG CTTTCAAGAA CTTGCAGGGG CTGTGGGGG TGCGGGCTCC
 12961 CACAGGCGAC CGCGCGACCG TGTCTAGCTT GCTGACGCC AACTCGCGCC TGTTGCTGCT
 13021 GCTAATAGCG CCCTTCACGG ACAGTGGCAG CGTGTCCCGG GACACATACC TAGGTCACTT
 13081 GCTGACACTG TACCGCGAGG CCATAGGTCA GGCGCATGTG GACGAGCATA CTTTCCAGGA
 13141 GATTACAAGT GTTAGCCCGC CGCTGGGCA GGAGGACACG GGCAGCCTGG AGGCAACCC
 13201 GAACTACCTG CTGACCAACC GGCGGCAAAA AATCCCCTCG TTGCACAGTT TAAACAGCGA
 13261 GGAGGAGCGC ATTTTGCCT ATGTGCAGCA GAGCGTGAGC CTTAACCTGA TGCGCGACGG
 13321 GGTAAACGCC AGCGTGGCGC TGGACATGAC CGCGCGAAC ATGGAACCGG GCATGTATGC
 13381 CTCAAACCCGG CCGTTTATCA ATCGCCTAAT GGACTACTTG CATCGCGGG CCGCCGTGAA
 13441 CCCCCGAGTAT TTCACCAATG CCATCTGAA CCCGCACTGG CTACCGCCCC CTGGTTCTA
 13501 CACCGGGGGA TTGAGGTG CCGAGGGTAA CGATGGATTC CTCTGGGACG ACATAGACGA
 13561 CAGCGTGTGTT TCCCCGCAAC CGCAGACCCCT GCTAGAGTTG CAACAAACGCG AGCAGGCAGA
 13621 GGCGGCGCTG CGAAAGGAAA GCTTCCGAG GCCAACCGAC TTGTCCGATC TAGGCCTGCG
 13681 GGCCCCCGGG TCAGATGCTA GTAGCCCATT TCCAAGCTTG ATAGGGTCTC TTACCAAGCAC
 13741 TCGCACCACC CGCCCGCGCC TGCTGGCGA GGAGGAGTAC CTAAACAACT CGCTGCTGCA
 13801 GCCCGAGCGC GAAAAGAACG TGCCCTCCGGC GTTTCCCAAC AACGGGATAG AGAGCCTAGT
 13861 GGACAAAGATG AGTAGATGGA AGACGTATGC GCAGGAGCAC AGGGATGTGC CCGGCCCCGG
 13921 CCCGCCCCACC CGTCGTCAAA GGCACGACCG TCAGCGGGGT CTGGTGTGGG AGGACGATGA
 13981 CTCGGCAGAC GACAGCAGCG TCTTGGATTT GGGAGGGAGT GGCAACCCGT TTGCACACCT
 14041 TCGCCCCCAGG CTGGGGAGAA TGTTTAAAAA AAAGCATGAT GCAAAATAAA AAACTCACCA
 14101 AGGCCATGGC ACCGAGCGTT GGTTTCTTG TATTCCCTT AGTATGCGC GCGCGGCGAT
 14161 GTATGAGGAA GGTCCCTCCCT CCTCCTACGA GAGCGTGGTG AGCGCGGCGC CAGTGGCGGC
 14221 GGCCTGGGT TCACCCCTCG ATGCTCCCCCT GGACCCGCGC TTCTGCTC CGCGGTACCT
 14281 GCGGCCCTACC GGGGGGAGAA ACAGCATCCG TTACTCTGAG TTGGCACCCCC TATTGACAC
 14341 CACCCGTGTG TACCTTGTGG ACAACAAGTC AACGGATGTG GCATCCCTGA ACTACCAGAA
 14401 CGACCACACG AACTTCTAA CCACGGTCAT TCAAAACAAAT GACTACAGCC CGGGGGAGGC
 14461 AAGCACACAG ACCATCAATC TTGACGACCG GTCGCACTGG GGCAGCGACCC TGAAAACCAT
 14521 CCTGCATACC AACATGCCAA ATGTGAACGA GTTCATGTTT ACCAATAAGT TTAAGGCGCG
 14581 GGTGATGGTG TCGCGCTCGC TTACTAAGGA CAAACAGGTG GAGCTGAAAT ACGAGTGGGT
 14641 GGAGTTCACTG CTGGCCGAGG GCAACTACTC CGAGACCATG ACCATAGACC TTATGAACAA
 14701 CGCGATCGTG GAGCACTACT TGAAAGTGGG CAGGCAGAAC GGGGTCTGG AAAGCGACAT
 14761 CGGGGTAAAG TTTGACACCC GCAACTTCAG ACTGGGGTTT GACCCAGTCA CTGGTCTTGT
 14821 CATGCCCTGGG GTATATACAA ACGAAGCCTT CCATCCAGAC ATCATTTGC TGCCAGGATG
 14881 CGGGGTGGAC TTCACCCACA GCCGCCCTGAG CAACTTGTG GGCATCCGCA AGCGGCAACC
 14941 CTTCCAGGAG GGCTTTAGGA TCACCTACGA TGACCTGGAG GGTGTAACA TTCCCGCACT
 15001 GTTGGATGTG GACGCCCTACC AGGCAAGCTT GAAAGATGAC ACCGAACAGG GCGGGGGTGG
 15061 CGCAGGCGGC GGCAACAAACA GTGGCAGCGG CGCGGAAGAG AACTCCAACG CGGCAGCTGC

FIG. 7F

15121 GGCAATGCAG CCGGTGGAGG ACATGAACGA TCATGCCATT CGCGGCGACA CCTTTGCCAC
15181 ACGGGCGGAG GAGAAGCGCG CTGAGGCCA GGAGCGGCCA GAAGCTGCCG CCCCCGCTGC
15241 GGAGGCTGCA CAACCCGAGG TCGAGAAGCC TCAGAAGAAA CCGGTGATTA AACCCCTGAC
15301 AGAGGACAGC AAGAAACGCA GTTACAACCT ATAAGCAAT GACAGCACCT TCACCCAGTA
15361 CCGCAGCTGG TACCTTGCA ACAACTACGG CGACCCCTCAG GCCGGGATCC GCTCATGGAC
15421 CCTGCTTGCA ACTCCTGACG TAACCTGCCG CTCGGAGCAG GTATACTGGT CGTGCCCCGA
15481 CATGATGCAA GACCCCGTGA CCTTCGCTC CACCGGCCAG ATCAGCAACT TTCCGGTGGT
15541 GGGGCCGAG CTGTTGCCG TGCACCTCAA GAGCTTCTAC AACGACCAGG CCGCTACTC
15601 CCAGCTCATC CGCCAGTTA CCTCTCTGAC CCACGTGTTA AATCGCTTTC CCGAGAACCA
15661 GATTTGGCG CGCCCGCCAG CCCCCACCAT CACCACCGTC AGTAAAACG TTCTGCTCT
15721 CACAGATCAC GGGACGCTAC CGCTGCGCAA CAGCATCGGA GGAGTCCAGC GAGTGACCAT
15781 TACTGACGCC AGACGCCGCA CCTGCCCTA CGTTTACAAG GCCCTGGGCA TAGTCTGCC
15841 GCGCGTCATA TCGAGGCCGA CTTTTGAGC AAACATGTCC ATCCTTATAT CGCCCAAGCAA
15901 TAACACAGGC TGGGGCCTGC GCTTCCAAAG CAAGATGTTT GGCGGGGCCA AGAAGCGCTC
15961 CGACCAACAC CCAGTGCAGC TGCGGGGCA CTACCGCGCG CCGTGGGGCG CGACAAACG
16021 CGGCCGCACT GGGCGCACCA CGTCGATGA CGCCATCGAC GCGGTGGTGG AGGAGGCGCG
16081 CAACTACACG CCCACGCCGC CGCCAGTGTG CACCGTGGAC GCGGCCATTG AGACCGTGGT
16141 GCGGGAGCC CGGGCCTACG CTAAAATGAA GAGACGGCGG AGGCGCGTAG CACGTCGCCA
16201 CGGCCGCCGA CGCCGCACTG CGCCCAACG CGCGCGGGCG GCGTGGCTTA ACCGCGCACG
16261 TCGCACCGGC CGACGGGGCG CCATGCGAGC CGCTCGAAGG CTGGCCGCGG GTATTGTCAC
16321 TGTGCCCCCCC AGGTCCAGGC GACGAGCGGC CGCCGCGAGCA GCGCGGGCCA TTAGTGCTAT
16381 GACTCAGGGT CGCAGGGCA ACGTGTACTG GGTGCGCGAC TCGTTAGCG GCCTGCGCGT
16441 GCCCGTGCAG ACCCGCCCCC CGCGCAACTA GATTGCAATA AAAAACTACT TAGACTCGTA
16501 CTGTTGTATG TATCCAGGG CGGGCGCGCG CATCGAAGCT ATGTCCAAGC GCAAAATCAA
16561 AGAAGAGATG CTCCAGGTCA TCGCGCCGGA GATCTATGGC CCCCCGAAGA AGGAAGAGCA
16621 GGATTACAAG CGCCGAAAGC TAAACGGGGT CAAAAAGAAA AAGAAAGATG ATGATGATGA
16681 TGAACTTGAC GACGAGGTGG AACTGTTGCA CGCGACCGCG CCCAGGCGAC GGGTACAGTG
16741 GAAAGGTGCA CGCGTAAGAC GTGTTTGCG ACCCGGCACC ACCGTAGTCT TTACGCCCG
16801 TGAGCGCTCC ACCCGCACCT ACAACGCCGT GTATGATGAG GTGTACGGCG ACGAGGACCT
16861 GCTTGAGCAG GCCAACGAGC GCCTCGGGGA GTTGCCTAC GGAAAGCGGC ATAAGGACAT
16921 GCTGGCGTTG CCGCTGGACG AGGGCAACCC AACACCTAGC CTAAAGCCCG TGACACTGCA
16981 GCAGGTGCTG CGCCGCGTTG CACCGTCCGA AGAAAAGCGC GGCCTAAAGC GCGAGTCTGG
17041 TGACTTGGCA CCCACCGTGC AGCTGATGGT ACCCAAGCGT CAGCGACTGG AAGATGTCTT
17101 GGAAAAAAATG ACCGTGGAGC CTGGGCTGGA GCGCGAGGTC CGCGTGCAGC CAATCAAGCA
17161 GGTGGCACCG GGACTGGCG TGAGACCGT GGACGTTCAAG ATACCCACCA CCAGTAGCAC
17221 TAGTATTGCC ACTGCCACAG AGGGCATGGA GACACAAAG TCCCCGGTTG CCTCGGGCGT
17281 GGCAGATGCC CGGGTGCAGG CGGGCGCTGC GGCGCGCTCC AAGACCTCTA CGGAGGTGCA
17341 AACGGACCCCG TGGATGTTTC GTGTTTCAGC CCCCCGGCGT CCGCGCCGTT CAAGGAAGTA
17401 CGGCGCCGCC AGCGCGCTAC TGCCCGAATA TGCCCTACAT CCTCCATCG CGCCTACCCC
17461 CGGCTATCGT GGCTACACCT ACCGCCCCAG AAGACGAGCA ACTACCCGAC GCCGAACAC
17521 CACTGGAACC CGCCGCGGCC GTCGCCGTCG CCAGCCCGTG CTGGCCCCGA TTTCCGTGCG
17581 CAGGGTGGCT CGCGAAGGAG GCAGGACCT GGTGCTGCCA ACAGCGCGCT ACCACCCAG

49/92

17641 CATCGTTAA AAGCCGGTCT TTGTGGTTCT TGCAGATATG GCCCTCACCT GCGCCCTCCG
 17701 TTTCCCGGTG CCGGGATTCC GAGGAAGAAT GCACCGTAGG AGGGGCATGG CCGGCCACGG
 17761 CCTGACGGGC GGCATGCGTC GTGCGCACCA CCGGCGGCGG CGCGCGTCGC ACCGTCGCAT
 17821 GCGCGGCGGT ATCCTGCCCG TCCTTATTCC ACTGATCGCC GCGGCGATTG CCGCCGTGCC
 17881 CGGAATTGCÀ TCCTGGCCT TGCAGGCAGA GAGACACTGA TTAAAAACAA GTTACATGTG
 17941 GAAAAATCAA AATAAAAGTC TGGACTCTCA CGCTCGCTTG GTCTGTAAAC TATTTTGTAG
 18001 AATGGAAGAC ATCAACTTGT CGTCAGTGGC CCCGCGACAC GGCTCGCGCC CGTTCATGGG
 18061 AAACCTGGCAA GATATCGGCA CCAGCAATAT GAGCGGTGGC GCCTTCAGCT GGGGCTCGCT
 18121 GTGGAGCGGC ATTAAAAATT TCGGTTCCGC CGTTAAGAAC TATGGCAGCA AAGCCTGGAA
 18181 CAGCAGCACA GGCCAGATGC TGAGGGACAA GTTGAAGAG CAAAATTTC CACAAAAGGT
 18241 GGTAGATGGC CTGGCCTCTG GCATTAGCGG GGTGGTGGAC CTGGCCAACC AGGCAGTGCA
 18301 AAATAAGATT AACAGTAAGC TTGATCCCCG CCCTCCCGTA GAGGAGCCTC CACCGGCCGT
 18361 GGAGACAGTG TCTCCAGAGG GGCCTGGCGA AAAGCGTCCG CGACCCGACA GGGAAAGAAC
 18421 TCTGGTGACG CAAATAGACG AGCCTCCCTC GTACGAGGAG GCACTAAAGC AAGGCTCGCC
 18481 CACCACCCGT CCCATCGCGC CCATGGCTAC CGGAGTGCTG GGCCAGCACA CACCGTAAC
 18541 GCTGGACCTG CCTCCCCCG CCGACACCCA GCAGAAACCT GTGCTGCCAG GCGCGTCCCG
 18601 CGTTGTTGTA ACCCGTCTTA GCGCGCGTC CCTGCGCCCG GCGCCAGCG GTCCCGCGATC
 18661 GTTGGGGCCC GTAGCCAGTG GCAACTGGCA AAGCACACTG AACAGCATCG TGGGTTTGGG
 18721 GGTGCAATCC CTGAAGGCC GACGATGCTT CTGATAGCTA ACGTGTCGTA TGTGTGTCAT
 18781 GTATGCGTCC ATGTCGCCGC CAGAGGAGCT GCTGAGCCCG CGCGCGCCCG CTTTCCAAGA
 18841 TGGCTACCCC TTGCGATGATG CCGCAGTGGT CTTACATGCA CATCTCGGGC CAGGACGCT
 18901 CGGAGTACCT GAGCCCCGGG CTGGTGCAGT TCGCCCGCGC CACCGAGACG TACTTCAGCC
 18961 TGAATAACAA GTTTAGAAAC CCCACGGTGG CGCCTACGCA CGACGTGACC ACAGACCGGT
 19021 CTCAGCGTTT GACGCTGCCG TTCACTCCCCG TGGACCGCGA GGATACTGCG TACTCGTACA
 19081 AGGCGCGTT CACCTAGCT GTGGGTGATA ACCGTGTGCT AGACATGGCT TCCACGTACT
 19141 TTGACATCCG CGGCGTGTG GACAGGGGCC CTACTTTAA GCCCTACTCT GGCACGTGCC
 19201 ACAACGCACT GGCCCCCAAG GGTGGCCCCA ACTCGTGCGA GTGGGAACAA AATGAAAATG
 19261 CACAAGTGGA TGCTCAAGAA CTTGACGAAG AGGAGAAATGA AGCCAATGAA GCTCAGGCC
 19321 GAGAACAGGA ACAAGCTAAG AAAACCCATG TATATGCCA GGCTCCACTG TCCGGAATAA
 19381 AAATAACTAA AGAAGGTCTA CAARTAGGAA CTGCCGACGC CACAGTAGCA GGTGCCGGCA
 19441 AAGAAAATTT CGCAGACAAA ACTTTCAAC CTGAACCACA AGTAGGGAGAA TCTCAATGGG
 19501 ACGAACGGGA TGCCACAGCA GCTGGTGGAA GGGTTCTTAA AAAGACAACT CCCATGAAAC
 19561 CCTGCTATGG CTCATACGCT AGACCCACCA ATTCCAACGG CGGACAGGGC GTTATGGTTG
 19621 AACAAAATGG TAAATTGGAA AGTCAAGTCG AAATGCAATT TTTTTCCACA TCCACAAATG
 19681 CCACAAATGA AGTTAACAA ATACAACCA CAGTTGTATT GTACAGCGAA GATGAAACA
 19741 TGGAAACTCC AGATACTCAT CTTTCTTATA AACCTAAAAT GGGGGATAAA AATGCCAAAG
 19801 TCATGCTTGG ACAACAAGCA ATGCCAAACA GACCAAATTA CATTGCTTTT AGAGACAATT
 19861 TTATTGGTCT CATGTATTAC AACAGCACAG GTAACATGGG TGTCCCTGCT GGTCAAGGCAT
 19921 CGCAGTTGAA CGCTGTTGTA GATTGCAAG ACAGAAAACAC AGAGCTGTCC TACCAAGCTT
 19981 TGCTTGATTC AATTGGCGAC AGAACAAAGAT ACTTTCAAT GTGGAATCAA GCTGTTGACA
 20041 GCTATGATCC AGATGTCAGA ATTATTGAGA ACCATGGAAC TGAGGATGAG TTGCAAATT
 20101 ATTGCTTCC TCTTGGTGGAA ATTGGGATTA CTGACACTTT TCAAGCTGTT AAAACAACTG

FIG. 7H

20161 CTGCTAACGG GGACCAAGGC AATACTACCT GGCAAAAAGA TTCAACATTT GCAGAACGCA
 20221 ATGAAATAGG GGTGGGAAAT AACTTGCCA TGGAAATTAA CCTGAATGCC AACCTATGGA
 20281 GAAATTCCT TTACTCCAAT ATTGCGCTGT ACCTGCCAGA CAAGCTAAAA TACAACCCCA
 20341 CCAATGTGGA AATATCTGAC AACCCCAACA CCTACGACTA CATGAACAAG CGAGTGGTGG
 20401 CTCCTGGCT TGTAGACTGC TACATTAACC TTGGGGCGCG CTGGTCTCTG GACTACATGG
 20461 ACAACGTTAA TCCCTTAAC CACCAACCGCA ATGCCGGCCT GCGTTACCGC TCCATGTTGT
 20521 TGGGAAACGG CCGCTACGTG CCCTTCACA TTCAGGTGCC CCAAAAGTTT TTTGCCATTA
 20581 AAAACCTCCT CCTCCTGCCA GGCTCATACA CATATGAATG GAACTTCAGG AAGGATGTTA
 20641 ACATGGTTCT GCAGAGCTCT CTGGGAAACG ACCTTAGAGT TGACGGGCT AGCATTAAAGT
 20701 TTGACAGCAT TTGCTTTAC GCCACCTTCT TCCCCATGGC CCACAAACAG GCCTCCACGC
 20761 TGGAAAGCCAT GCTCAGAAAT GACACCAACG ACCAGTCCTT TAATGACTAC CTTTCCGCCG
 20821 CCAACATGCT ATATCCATA CCCGCCAACG CCACCAACGT GCCCATCTCC ATCCCATCGC
 20881 GCAACTGGGC AGCATTTCGC GGTTGGGCCT TCACACGCTT GAAGACAAAG GAAACCCCTT
 20941 CCCTGGGATC AGGCTACGAC CCTTACTACA CCTACTCTGG CTCCATACCA TACCTTGACG
 21001 GAACCTTCTA TCTTAATCAC ACCTTTAAGA AGGTGGCCAT TACTTTTGAC TCTTCTGTTA
 21061 GCTGGCCGGG CAACGACCGC CTGCTTACTC CCAATGAGTT TGAGATTAAG CGCTCAGTTG
 21121 ACGGGGAGGG CTATAACGTA GCTCAGTGCA ACATGACAAA GGACTGGTTC CTAGTGCAGA
 21181 TGTTGGCCAA CTACAATATT GGCTACCAAGG GCTTCTACAT TCCAGAAAGC TACAAAGACC
 21241 GCATGTACTC GTTCTTCAGA AACTTCCAGC CCATGAGCCG GCAAGTGGTG GACGATACTA
 21301 AATACAAAGA TTATCAGCAG GTTGAATTAA TCCACCAGCA TAACAACCTCA GGCTTCGTAG
 21361 GCTACCTCGC TCCCACCATG CGCGAGGGAC AAGCTTACCC CGCTAATGTT CCCTACCCAC
 21421 TAATAGGCAA AACCGCGGTT GATAGTATTA CCCAGAAAAA GTTCTTTGC GACCGCACCC
 21481 TGTTGGCGCAT CCCCTTCTCC AGTAACTTA TGTCATGGG TGCGCTCACA GACCTGGGCC
 21541 AAAACCTTCT CTACGAAAC TCCGCCACG CGCTAGACAT GACCTTTGAG GTGGATCCCA
 21601 TGGACGAGCC CACCCCTCTT TATGTTTGT TTGAAGTCTT TGACGTGGTC CGTGTGCACC
 21661 AGCCGCACCG CGGCGTCATC GAGACCGTGT ACCTGCCAC GCCCTCTCG GCCGGCAACCG
 21721 CCACAACATA AAGAAGCAAG CAACATCAAC AACAGCTGCC GCCATGGGCT CCAGTGAGCA
 21781 GGAAGTGAAG GCCATTGTCA AAGATCTTGG TTGTGGGCCA TATTTTTGG GCACCTATGA
 21841 CAAGCGCTTC CCAGGCTTGT TTTCACCAAA CAAGCTGCC TGCGCCATAG TTAACACGGC
 21901 CGGTCGCGAG ACTGGGGCG TACACTGGAT GGCTTTGCC TGGAACCCGC GCTAAAAAAC
 21961 ATGCTACCTC TTGAGCCCT TTGGCTTTTC TGACCAACGT CTCAAGCAGG TTTACCAGTT
 22021 TGAGTACGAG TCACTCTGC GCCGTAGCGC CATTGCCTCT TCCCCCGACC GCTGTATAAC
 22081 GCTGGAAAAG TCCACCCAAA GCGTGCAGGG GCCCAACTCG GCCGCCTGTG GCCTATTCTG
 22141 CTGCATGTTT CTCCACGCCCT TTGCCAACTG GCCCCAAACT CCCATGGATC ACAACCCAC
 22201 CATGAACCTT ATTACCGGGG TACCCAACTC CATGCTTAAC AGTCCCCAGG TACAGCCAC
 22261 CCTGCGCCGC AACCAAGGAAC AGCTCTACAG CTTCCCTGGAG CGCCACTCGC CCTACTTCCG
 22321 CAGCCACAGT GCGAAATTA GGAGCGCCAC TTCTTTTGT CACTGAAAA ACATGTAAAA
 22381 ATAATGTACT AGGAGACACT TTCAATAAG GCAAATGTT TTATTTGTAC ACTCTCGGGT
 22441 GATTATTTAC CCCCCACCTT GCGCTCTGGC CGGTTTAAAATCAAAGGGG TTCTGCCGCC
 22501 CATCGCTATG CGCCACTGGC AGGGACACGT TGCGATACTG GTGTTAGTG CTCCACTTAA
 22561 ACTCAGGCAC AACCATCCGC GGCAGCTCGG TGAAGTTTC ACTCCACAGG CTGCGCACCA
 22621 TCACCAACGC GTTTAGCAGG TCGGGCGCCG ATATCTGAA GTCGCAGTTG GGGCCTCCGC

22681 CCTGCGCGCG CGAGTTGCGA TACACAGGGT TACAGCACTG GAACACTATC AGGCCGGGT
 22741 GGTGCACGCT GGCCAGCACG CTCTTGTGCG AGATCAGATC CGCGTCCAGG TCCTCCGCGT
 22801 TGCTCAGGGC GAACGGAGTC AACTTGGTA GCTGCCTTCC CAAAAAGGGT GCATGCCAG
 22861 GCTTGAGTT GCACTCGCAC CGTAGTGGCA TCAGAAGGTG ACCGTGCCA GTCTGGCGT
 22921 TAGGATACAG CGCCTGCATG AAAGCCTTGA TCTGCTTAA AGCCACCTGA GCCTTGGCG
 22981 CTTCAAGAGAA GAACATGCCG CAAGACTTGC CGGAAAACGT ATTGGCCGGA CAGGCCGCGT
 23041 CATGCACGCA GCACCTTGC CGGGTGTGG AGATCTGCAC CACATTTCGG CCCCACCGGT
 23101 TCTTCACGAT CTTGGCCTTG CTAGACTGCT CCTTCAGCGC GCGCTGCCG TTTTCGCTCG
 23161 TCACATCCAT TTCAATCAG TGCTCCTTAT TTATCATAAT GCTCCCGTGT AGACACTTAA
 23221 GCTCGCCTTC GATCTCAGCG CAGCGGTGCA GCCACAAACGC GCAGGCCGTG GGCTCGTGGT
 23281 GCTTGAGGT TACCTCTGCA AACGACTGCA GGTACGCCGT CAGGAATCGC CCCATCATCG
 23341 TCACAAAGGT CTTGTTGCTG GTGAAGGTCA GCTGCAACCC GCGGTGCTCC TCGTTTAGCC
 23401 AGGTCTTGC A TACGGCCGCC AGACCTTCCA CTTGGTCAGG CAGTAGCTTG AAGTTTGCCT
 23461 TTAGATCGTT ATCCACGTGG TACTTGTCCA TCAACGCGC CGCAGCCTCC ATGCCCTTCT
 23521 CCCACGCAGA CACGATCGGC AGGCTCAGCG GGTTTATCAC CGTGCTTCA CTTTCCGCTT
 23581 CACTGGACTC TTCCCTTTCC TCTTGCACTCC GCATACCCCG CGCCACTGGG TCGTCTTCAT
 23641 TCAGCCGCCG CACCGTGC CG TTACCTCCCT TGCCGTGCTT GATTAGCACC GGTGGGTTGC
 23701 TGAAACCCAC CATTGTAGC GCCACATCTT CTCTTTCTC CTCGCTGTCC ACGATCACCT
 23761 CTGGGGATGG CGGGCGCTCG GGCTTGGAG AGGGCGCTT CTTTTCTTT TTGGACGCAA
 23821 TGGCCAAATC CGCCGTCGAG GTCGATGGCC GCGGGCTGGG TGTGCGCGGC ACCAGCGCAT
 23881 CTTGTGACGA GTCTTCTTCG TCCTCGGACT CGAGACGCCG CCTCAGCCGC TTTTTTGGGG
 23941 GCGCGCGGGG AGGCGGCCGC GACGGCGACG GGGACGAGAC GTCCTCCATG GTTGGTGGAC
 24001 GTCGCGCCGC ACCCGTCCG CGCTGGGGG TGTTTTCGCG CTGCTCCCTCT TCCCGACTGG
 24061 CCATTTCCCTT CTCCTATAGG CAGAAAAAGA TCATGGAGTC AGTCGAGAAG GAGGACAGCC
 24121 TAACCGCCCC CTTTGAGTTG GCCACCACCG CCTCCACCGA TGCCGCCAAC GCGCCTACCA
 24181 CCTTCCCCGT CGAGGCACCC CCGCTTGAGG AGGAGGAAGT GATTATCGAG CAGGACCCAG
 24241 GTTTGTAAAG CGAAGACGAC GAAGATCGCT CAGTACCAAC AGAGGATAAA AAGCAAGACC
 24301 AGGACGACGC AGAGGAAAC GAGGAACAAG TCGGGCGGGG GGACCAAAGG CATGGCGACT
 24361 ACCTAGATGT GGGAGACGAC GTGCTGTGAG AGCATCTGCA GCGCAGTGC GCCATTATCT
 24421 GCGACGCGTT GCAAGAGCGC AGCGATGTGC CCCTCGCCAT AGCGGATGTC AGCCTTGCT
 24481 ACGAACGCCA CCTGTTCTCA CGCGCGTAC CCCCCAAACG CCAAGAAAAC GGCACATGCG
 24541 AGCCAACCC GCGCCTCAAC TTCTACCCCG TATTTGCCGT GCCAGAGGTG CTTGCCACCT
 24601 ATCACATCTT TTTCCAAAAC TGCAAGATAC CCCTATCCTG CCGTCCAAC CGCAGCCGAG
 24661 CGGACAAGCA GCTGGCCTTG CGGCAGGGCG CTGTCATACC TGATATCGCC TCGCTCGACG
 24721 AAGTGCCAAA AATCTTGAG GGTCTTGGAC GCGACGAGAA GCGCGCGCA AACGCTCTGC
 24781 AACAAAGAAAA CAGCGAAAAT GAAAGTCACT GTGGAGTGCT GGTGGAACCTT GAGGGTGA
 24841 ACGCGCGCT AGCCGTGCTG AAACCGAGCA TCGAGGTACAC CCACCTTGCC TACCCGGCAC
 24901 TTAACCTTACCC CCCCCAAGGTT ATGAGCACAG TCATGAGCGA GCTGATCGTG CGCCGTGCA
 24961 GACCCCTGGA GAGGGATGCA AACTTGCAAG AACAAACCGA GGAGGGCTA CCCGAGTTG
 25021 GCGATGAGCA GCTGGCGCGC TGGCTTGAGA CGCGCGAGCC TGCCGACTTG GAGGAGCGAC
 25081 GCAAGCTAAT GATGGCCGCA GTGCTTGTGAGCT TGAGTGCATG CAGCGGTTCT
 25141 TTGCTGACCC GGAGATGCAG CGCAAGCTAG AGGAAACGTT GCACTACACC TTTCGCCAGG

52/92

25201 GCTACGTGCG CCAGGCCTGC AAAATTCCA ACGTGGAGCT CTGCAACCTG GTCTCCTACC
 25261 TTGGAATTTT GCACGAAAAC CGCCTGGGC AAAACGTGCT TCATTCCACG CTCAAGGGCG
 25321 AGGCGCGCCG CGACTACGTG CGCGACTGGG TTTACTTATT TCTGTGCTAC ACCTGGAAA
 25381 CGGCCATGGG CGTGTGGCAG CAGTGCCTGG AGGAGCGCAA CCTGAAGGAG CTGCAGAAGC
 25441 TGCTAAAGCA AAACTTGAAG GACCTATGGA CGGCCTTCAA CGAGCGCTCC GTGGCCGCGC
 25501 ACCTGGCGGA CATTATCTC CCCGAACGCC TGCTTAAAC CCTGCAACAG GGTCTGCCAG
 25561 ACTTCACCAAG TCAAAGCATG TTGCAAAACT TTAGGAACCTT TATCCTAGAG CGTTCAAGGAA
 25621 TTCTGCCCGC CACCTGCTGT GCGCTTCCTA GCGACTTTGT GCCCATTAAAG TACCGTGAAT
 25681 GCCCTCCGCC GCTTGGGGT CACTGCTACC TTCTGCAGCT AGCCAACATAC CTTGCCTACC
 25741 ACTCCGACAT CATGGAAGAC GTGAGCGGTG ACGGCCACT GGAGTGTAC TGTCGCTGCA
 25801 ACCTATGCAC CCCGCACCGC TCCCTGGTCT GCAATTCAACA ACTGCTTAGC GAAAGTCAAA
 25861 TTATCGGTAC CTTTGAGCTG CAGGGTCCCT CGCCTGACGA AAAGTCCGCG GCTCCGGGGT
 25921 TGAAACTCAC TCCGGGGCTG TGGACGTCGG CTTACCTTCG CAAATTGTA CCTGAGGACT
 25981 ACCACGCCCA CGAGATTAGG TTCTACGAAG ACCAATCCCG CCCGCCAAAT GCGGAGCTTA
 26041 CGCCCTGCGT CATTACCCAG GCCCACATCC TTGGCCAATT GCAAGCCATT AACAAAGGCC
 26101 GCCAAGAGTT TCTGCTACGA AAGGGACGGG GGGTTTACTT GGACCCCCAG TCCGGCGAGG
 26161 AGCTCAACCC AATCCCCCG CGCCCGCAGC CCTATCAGCA GCCGCGGGCC CTTGCTTCCC
 26221 AGGATGGCAC CCAAAAAGAA GCTGCAGCTG CGGCCGCCGC CACCCACGGG CGAGGAGGAA
 26281 TACTGGGACA GTCAGGCAGA GGAGGTTTTG GACGAGGAGG AGGAGATGAT GGAAGACTGG
 26341 GACAGCCTAG ACGAGGAAGC TTCCGAGGCC GAAGAGGTGT CAGACGAAAC ACCGTCACCC
 26401 TCGTCGCAT TCCCCTCGCC GGCGCCCGAG AAATCGGCAA CGTTCCCGAG CATTGCTACA
 26461 ACCTCCGCTC CTCAGGCCCG GCCGGCACTG CCCGTTGCC GACCCACCG TAGATGGGAC
 26521 ACCACTGGAA CCAGGGCCGG TAAGTCTAAG CAGCCGCCGC CGTTAGCCA AGAGCAACAA
 26581 CAGCGCCAAG GCTACCGCTC GTGGCGCGTG CACAAGAACG CCATAGTTGC TTGCTTGCAA
 26641 GACTGTGGGG GCAACATCTC CTTCGCCCGC CGCTTTCTTC TCTACCATCA CGCGTGGCC
 26701 TTCCCCCGTA ACATCTGCA TTACTACCGT CATTCTCTACA GCCCCTACTG CACCGGCCGG
 26761 AGCGGCAGCA ACAGCAGCGG CCACGCAGAA GCAAAGGCCA CGGGATAGCA AGACTCTGAC
 26821 AAAGCCCAAG AAATCCACAG CGGCGGCAGC AGCAGGAGGA GGAGCACTGC GTCTGGCGCC
 26881 CAACGAACCC GTATCGACCC GCGAGCTTAG AACAGGATT TTCCCCACTC TGTATGCTAT
 26941 ATTCAACAG AGCAGGGCC AAGAACAAAGA GCTGAAAATA AAAAACAGGT CTCTGCCTC
 27001 CCTCACCGC AGCTGCTGT ATCACAAAAG CGAAGATCAG CTTCGCGCA CGCTGGAAGA
 27061 CGCGGAGGCT CTCTTCAGCA AATACTGCGC GCTGACTCTT AAGGACTAGT TTGCGCCCT
 27121 TTCTCAAATT TAAGCGGAA AACTACGTCA TCTCCAGCGG CCACACCCGG CGCCAGCACC
 27181 TGTCGTCAGC GCCATTATGA GCAAGGAAAT TCCCACGCC TACATGTGGA GTTACCGAGCC
 27241 ACAAAATGGGA CTTCGGGCTG GAGCTGCCA AGACTACTCA ACCCGAATAA ACTACATGAG
 27301 CGCGGGACCC CACATGATAT CCCGGGTCAA CGGAATCCCG GCCCCACCGAA ACCGAATTCT
 27361 CCTCGAACAG GCGGCTATTA CCACCCACCC TCGTAATAAC CTTAACCCCC GTAGTTGGCC
 27421 CGCTGCCCTG GTGTACCAAGG AAAGTCCCGC TCCCACCACT GTGGTACTTC CCAGAGACGC
 27481 CCAGGCCGAA GTTCAGATGA CTAACTCAGG GGCGCAGCTT GCAGGGCGGCT TTGTCACAG
 27541 GGTGCGGTGCG CCCGGGCAGG GTATAACTCA CCTGAAAATC AGAGGGCGAG GTATTCAAGCT
 27601 CAACGACGAG TCGGTGAGCT CCTCTCTTGG TCTCCGTCCG GACGGGACAT TTGAGATCGG
 27661 CGGCCTGGC CGCTCTTCAT TTACGCCCG TCAGGCGATC CTAACCTGTC AGACCTCGTC

FIG. 7K

53/92

27721 CTCGGAGCCG CGCTCCGGAG GCATGGAAAC TCTACAATT ATTGAGGAGT TCGTGCCCTTC
 27781 GGTACTTC AACCCCTTT CTGGACCTCC CGGCCACTAC CGGGACCAGT TTATCCCCAA
 27841 CTTGACGCG GTAAAAGACT CGGCGGACGG CTACGACTGA ATGACCACTG GAGAGGGCAGA
 27901 GCAACTGCGC CTGACACACC TCGACCCTG CCGCCGCCAC AAGTGCTTTG CCCGCGGCTC
 27961 CGGTGAGTTT TGTTACTTGT AATTGCCGA AGAGCATATC GAGGGCCCGG CGCACGGCGT
 28021 CCGGCTCACCC ACCCAGGTAG AGCTTACACG TAGCCTGATT CGGGAGTTA CCAAGGCC
 28081 CCTGCTAGTG GAGCGGGAGC GGGGTCCTG TGTTCTGACC GTGGTTTGCA ACTGTCCTAA
 28141 CCCTGGATTAA CATCAAGATC TTTGTTGTCA TCTCTGTGCT GAGTATAATA AATACAGAAA
 28201 TTAGAATCTA CTGGGGCTCC TGTGCCATC CTGTGAACGC CACCGTTTT ACCCACCCAA
 28261 AGCAGACCAA AGCAAACCTC ACCTCCGGTT TGCACAAGCG GGCCAATAAG TACCTTACCT
 28321 GGTACTTTAA CGGCTCTTC TTTGTAATT ACAACAGTTT CCAGCGAGAC GAAGTAAGTT
 28381 TGCCACACAA CCTTCTCGGC TTCAACTACA CGGTCAAGAA AAACACCACC ACCACCCCTCC
 28441 TCACCTGCCG GGAACGTACG AGTGCAC TCGTTGCTGC GCCCACACCT ACAGCCTGAG
 28501 CGTAACCAGA CATTACTCCC ATTTCCCAA AACAGGAGGT GAGCTCAACT CCCGGAAC
 28561 AGGTCAAAAA AGCATTTCG GGGGTGCTGG GATTTTTAA TTAAGTATAT GAGCAATTCA
 28621 AGTAACTCTA CAAGCTTGTCA TAATTTTCT GGAATTGGGG TCGGGGTTAT CCTTACTCTT
 28681 GTAATTCTGT TTATTCTTAT ACTAGCACTT CTGTGCCTTA GGGTTGCCGC CTGCTGCACG
 28741 CACGTTTGTA CCTATTGTCA GCTTTTAAA CGCTGGGGGC GACATCCAAG ATGAGGTACA
 28801 TGATTTAGG CTTGCTGCC CTTGCGCAG TCTGCAGCG TGCCAAAAAG GTTGAGTTA
 28861 AGGAACCAGC TTGCAATGTT ACATTTAAAT CAGAAGCTAA TGAATGCACT ACTCTTATAA
 28921 AATGCACCAC AGAACATGAA AAGCTTATTA TTGCGCACAA AGACAAAATT GGCAAGTATG
 28981 CTGTATATGC TATTTGGCAG CCAGGTGACA CTAACGACTA TAATGTCACA GTCTTCCAAG
 29041 GTGAAAATCG TAAAACTTT ATGATAAAAT TTCCATTGTA TGAAATGTGC GATATTACCA
 29101 TGTACATGAG CAAACAGTAC AAGTTGTGGC CCCCCACAAA GTGTTAGAG AACACTGGCA
 29161 CCTTTGTTCA CACCGCTCTG CTTATTACAG CGCTTGCTTT GGTATGTACC TTACTTTATC
 29221 TCAAATACAA AAGCAGACGC AGTTTATTG ATGAAAAGAA AATGCCCTGA TTTTCCGC
 29281 GCTTGTATTTC CCCTGGACAA TTTACTCTAT GTGGGATATG CGCCAGGCAG GAAAGATTAT
 29341 ACCCACAACC TTCAAATCAA ACTTTCTGG ACGTTAGCGC CTGACTCTG CCAGCGCCTG
 29401 CACTGCAAAT TTGATCAAAC CCAGCTTCAG CTTGCCTGCT CCAGAGATGA CCGGCTCAAC
 29461 CATCGCGCC ACAACGGACT ATCGCAACAC CACTGCTACC GGACTAAAAT CTGCCCTAAA
 29521 TTACCCCCAA GTTCATGCTT TTGTCATGAA CTGGGCGAGC TTGGGCATGT GGTGGTTTC
 29581 CATAGCGCTT ATGTTTGTGTT GCCTTATTAT TATGTTGGCTT ATTTGTTGCC TAAAGCGCAG
 29641 ACGGCCAGA CCCCCCATCT ATAGGCCTAT CATTGTCGCTC AACCCACACAA ATGAAAAAAT
 29701 TCATAGATTG GACGGCTCA AACCATGTT TCTTCTTTA CAGTATGATT AAATGAGACA
 29761 TGATTCCTCG AGTCCTTATA TTATTGACCC TTGTTGCGCT TTTCTGTGCG TGCTCTACAT
 29821 TGGCTGCGGT CGCTCACATC GAAGTAGATT GCATCCCACC TTTCACAGTT TACCTGCTTT
 29881 ACGGATTGTG CACCCATTATC CTCATCTGCA GCCTCGTAC TGTAGTCATC GCCTTCATTC
 29941 AGTCATTGA CTGGATTGTG GTGCGCATTG CGTACCTTAG GCACCATCCG CAATACAGAG
 30001 ACAGGACTAT AGCTGATCTT CTCAGAATTG TTTAATTATG AAACGGATTG TCACTTTG
 30061 TTGCTGATT TTCTGCCGC TACCTGTGCT TTGCTCCCAA ACCTCAGCGC CTCCAAAAG
 30121 ACATATTTC TGCAGATTCA CTCAAATATG GAACATTCCC AGCTGCTACA ACAAAACAGAG
 30181 CGATTTGTCA GAAGCCTGGT TATACGCCAT CATCTGTGTC ATGGTTTTT GCAGTACCAT

FIG. 7L

54/92

30241 TTTTGCCTA GCCATATACC CATACTTGA CATTGGTTGG AATGCCATAG ATGCCATGAA
 30301 CCACCCCTACT TTCCCCAGCGC CCAATGTCAT ACCACTGCAA CAGGTTATTG CCCCCAATCAA
 30361 TCAGCCCTCGC CCCCCCTCTC CCACCCCCAC TGAGATTAGC TACTTTAATT TGACAGGTGG
 30421 AGATGACTGA ATCTCTAGAT CTAGAATTGG ATGGAATTAA CACCGAACAG CGCCTACTAG
 30481 AAAGGCGCAA GGCGCGTCC GAGCGAGAAC GCCTAAAACA AGAAGTTGAA GACATGGTTA
 30541 ACCTGCACCA GTGTAAAAGA GGTATCTTT GTGTGGTCAA GCAGGCCAAA CTTACCTACG
 30601 AAAAAACAC TACCGGCAAC CGCCTTAGCT ACAAGCTACC CACCCAGCGC CAAAAACTGG
 30661 TGCTTATGGT GGGAGAAAAA CCTATCACCG TCACCCAGCA CTCGGCAGAA ACAGAAGGCT
 30721 GCCTGCACCT CCCCTATCAG GGTCCAGAGG ACCTCTGCAC TCTTATTAAA ACCATGTGTG
 30781 GCATTAGAGA TCTTATTCCA TTCAACTAAC AATAAACACA CAATAAATTAA CTTACTTTAA
 30841 ATCAGTCAGC AAATCTTGT CCAGCTTATT CAGCATCACC TCCTTCCCT CCTCCCAACT
 30901 CTGGTATTTG AGCAGCCTT TAGCTGCAGA CTTTCTCCAA AGTCTAAATG GGATGTCAA
 30961 TTCCCTCATGT TCTTGTCCCT CCGCACCCAC TATCTTCATA TTGTTGAGA TGAAACGCGC
 31021 CAGACCGTCT GAAGACACCT TCAACCCCTGT GTACCCATAT GACACGGAAA CCGGCCCTCC
 31081 AACTGTGCCT TTCCCTTACCC CTCCCTTGT GTCGCCAAAT GGGTTCCAAG AAAGTCCCC
 31141 CGGAGTGCTT TCTTGCCTC TTTCAGAACC TTGTTTACCC TCACACGGCA TCCTTGCCT
 31201 AAAAAATGGGC AGCGGCCTGT CCCTGGATCA GGCAGGCAAC CTTACATCAA ATACAATCAC
 31261 TGTTTCTCAA CCGCTAAAAA AAACAAAGTC CAATATAACT TTGAAACAT CCGCGCCCC
 31321 TACAGTCAGC TCAGGCGCCC TAACCATGGC CACAACCTCG CCTTGGTGG TCTCTGACAA
 31381 CACTCTTACC ATGCAATCAC AAGCACCGCT AACCGTGCAA GACTCAAAAC TTAGCATTGC
 31441 TACCAAAGAG CCACTTACAG TGTTAGATGG AAAACTGGCC CTGCAGACAT CAGCCCC
 31501 CTCTGCCACT GATAACAACG CCCTCACTAT CACTGCCTCA CCTCCTCTTA CTACTGC
 31561 TGGTAGTCTG GCTGTTACCA TGGAAAACCC ACTTTACAAC ACAATGGAA AACTTGGCT
 31621 CAAAATTGGC GGTCTTTGC AAGTGGCCAC CGACTCACAT GCACTAACAC TAGGTACTGG
 31681 TCAGGGGGTT GCAGTTCATA ACAATTGCT ACATACAAA GTTACAGGCG CAATAGGGTT
 31741 TGATACATCT GGCAACATGG AACTAAAAAC TGGAGATGGC CTCTATGTGG ATAGGCCGG
 31801 TCCTAACCAA AAACTACATA TTAATCTAAA TACCACAAA GGCCTGCTT TTGACAACAC
 31861 CGCAATAACA ATTAACGCTG GAAAAGGGTT GGAATTGAA ACAGACTCCT CAAACGGAAA
 31921 TCCCATAAAA AAAAAATTG GATCAGGCAT ACAATATAAT ACCAATGGAG CTATGGTGC
 31981 AAAACTTGGA ACAGGCCCTCA GTTTGACAG CTCCGGAGCC ATAACAATGG GCAGCATAAA
 32041 CAATGACAGA CTTACTCTT GGACAAACACC AGACCCATCC CAAATTGCA GAATTGCTTC
 32101 AGATAAAGAC TGCAAGCTAA CTCTGGCCT AACAAAATGT GGCAGTCAAA TTTTGGGCAC
 32161 TGGTTCAGCT TTGGCAGTAT CAGGTAATAT GGCCTCCATC AATGGAACTC TAAGCAGTGT
 32221 AACTTGGTT CTTAGATTTG ATGACAACGG AGTGCTTATG TCAAATTCTAT CACTGGACAA
 32281 ACAGTATTGG AACTTTAGAA ACGGGGACTC CACTAACGGT CAACCATAACA CTTATGCTGT
 32341 TGGTTTATG CAAACCTAA AAGCTTACCC AAAACTCAA AGTAAAATG CAAAAAGTAA
 32401 TATTGTTAGC CAGGTGTATC TTAATGGTGA CAAGTCTAAA CCATTGCTT TTACTATTAC
 32461 GCTAAATGGGA ACAGATGAAA CCAACCAAGT AAGCAAATAC TCAATATCAT TCAGTTGGTC
 32521 CTGGAACAGT GGACAATACA CTAATGACAA ATTTGCCACC AATTCTATA CCTTCTCCTA
 32581 CATTGCCAG GAATAAAAGAA TCGTGAACCT GTTGCATGTT ATGTTCAAC GTGTTTATT
 32641 TTCAATTGCA GAAAATTCA AGTCATTTT CATTCAAGTAG TATAGCCCCA CCACCACATA
 32701 GCTTATACTA ATCACCGTAC CTTAATCAA CTCACAGAAC CCTAGTATTG AACCTGCCAC

FIG. 7M

32761 CTCCCTCCC ACACACAGAG TACACAGTCC TTTCTCCCCG GCTGGCCTTA AACAGCATCA
 32821 TATCATGGGT AACAGACATA TTCTTAGGTG TTATATTCCA CACGGCTCTCC TGTCGAGCCA
 32881 AACGCTCATC AGTGATGTTA ATAAACTCCC CGGGCAGCTC GCTTAAGTTC ATGTCGCTGT
 32941 CCAGCTGCTG AGCCACAGGC TGCTGTCAA CTTGCGGTG CTCAACGGGC GGCAGGGAG
 33001 AAGTCCACGC CTACATGGGG GTAGAGTCAT AATCGTGCAT CAGGATAGGG CGGTGGTGCT
 33061 GCAGCAGCGC GCGAATAAAC TGCTGCCGC CGCGCTCGT CCTGCAGGAA TACAACATGG
 33121 CAGTGGTCTC CTCAGCGATG ATTCCGACCG CCCGCAGCAT AAGGCCTT GTCCTCCGGG
 33181 CACAGCAGCG CACCCGTGTC TCACTTAAGT CAGCACAGTA ACTGCAGCAC AGTACCCACAA
 33241 TATTGTTAA AATCCCACAG TGCAAGGCG TGATATCCAAA GCTCATGGCG GGGACCACAG
 33301 AACCCACGTG GCCATCATAAC CACAAGCGCA GGTAGATTAA GTGGCGACCC CTCATAAAACA
 33361 CGCTGGACAT AAACATTACC TCTTTGGCA TGTTGTAATT CACCACCTCC CGGTACCATTA
 33421 TAAACCTCTG ATTAAACATG CGGCCATCCA CCACCATCCT AAACCAAGCTG GCCAAAACCT
 33481 GCCCGCCGGC TATGCACTGC AGGGAACCGG GACTGGAACA ATGACAGTGG AGAGCCCAGG
 33541 ACTCGTAACC ATGGATCATC ATGCTCGTCA TGATATCAAT GTTGGCACAA CACAGGCACA
 33601 CGTGCATAACA CTTCCCTCAGG ATTACAAGCT CCTCCCGCT CAGAACCCATA TCCCAGGGAA
 33661 CAACCCATTC CTGAATCAGC GTAAATCCC CACTGCAGGG AAGACCTCGC ACGTAACTCA
 33721 CGTTGTGCTAT GTCAAAGTG TTACATTCCGG GCAGCAGCGG ATGATCCTCC AGTATGGTAG
 33781 CGCGTGTCTC TGTCTCAAAA GGAGGTAGGC GATCCCTACT GTACGGAGTG CGCCGAGACA
 33841 ACCGAGATCG TGTGGTCTGT AGTGTCACTGC CAAATGGAAC CGCGGACGTA GTCATATTTC
 33901 CTGAAGCAAA ACCAGGTGCG GGCCTGACAA ACAGATCTGC GTCTCCGGTC TCGTCGTTA
 33961 GCTCGCTCTG TGTAGTAGTT GTAGTATATC CACTCTCTCA AAGCATCCAG GCGCCCCCTG
 34021 GCTTCGGTT CTATGTAAAC TCCCTCATGC CGCGCTGCC TGATAACATC CACCCACCGCA
 34081 GAATAAGCCA CACCCAGCCA ACCTACACAT TCGTTCTGGC AGTCACACAC GGGAGGAGCG
 34141 GGAAGAGCTG GAAGAACCAT GTTTTTTTTT TTTATTCCAA AAGATTATCC AAAACCTCAA
 34201 AATGAAGATC TATTAAGTGA ACGCGCTCCC CTCCGGTGGC GTGGTCAAAC TCTACAGCCA
 34261 AAGAACAGAT AATGGCATTG GTAAGATGTT GCACAATGGC TTCCAAAAGG CAAACTGCC
 34321 TCACGTCCAA GTGGACGTA AGGCTAAACC CTTCAGGGTG AATCTCCTCT ATAAACATTC
 34381 CAGCACCTTC AACCATGCCA AAATAATTTC CATCTGCCA CCTTATCAAT ATGTCTCTAA
 34441 GCAAATCCC AATATTAAGT CCGGCCATTG TAAAAATCTG CTCCAGAGCG CCCTCCACCT
 34501 TCAGCCTCAA GCAGCGAATC ATGATTGCAA AAATTCAAGT TCCTCACAGA CCTGTATAAG
 34561 ATTCAAAAGC GGAACATTAA CAAAAATACC GCGATCCCCTG AGGTCCCTTC GCAGGGCCAG
 34621 CTGAACATAA TCGTGCAGGT CTGCACGGAC CAGCGGGCC ACTTCCCCGC CAGGAACCAT
 34681 GACAAAAGAA CCCACACTGA TTATGACACG CATACTCGGA GCTATGCTAA CCAGCGTAGC
 34741 CCCGATGTAA GCTTGTGCA TGGCGGCCGA TATAAAATGC AAGGTACTGC TCAAAAATC
 34801 AGGCAAAGCC TCGCGAAAAA AAGCAAGCAC ATCGTAGTCA TGCTCATGCA GATAAAGGCA
 34861 GGTAAGTTC GGAACCCACCA CAGAAAAAGA CACCATTTC CTCTCAAACA TGTCGCGGG
 34921 TTCCTGCATA AACACAAAAT AAAATAACAA AAAAAAAA ACATTTAAC ATTAGAAGCC
 34981 TGTNTTACAA CAGGAAAAAC AACCTTATA AGCATAAGAC GGACTACGGC CATGCCGGC
 35041 TGACCGTAA AAAACTGGTC ACCGTGATTA AAAAGCACCA CCGACAGTTC CTCGGTCATG
 35101 TCCGGAGTCA TAATGTAAGA CTCGGTAAAC ACATCAGGTT GGTTAACATC GGTCAGTGCT
 35161 AAAAAGCGAC CGAAATAGCC CGGGGGATA CATAACCGCA GGCAGGAGA CAACATTACA
 35221 GCCCCCCATAG GAGGTATAAC AAAATTAAATA GGAGAGAAAA ACACATAAAC ACCTGAAAAA

56/92

35281 CCCTCCTGCC TAGGCAAAAT AGCACCCCTCC CGCTCCAGAA CAACATACAG CGCTTCCACA
35341 GCGGCAGCCA TAACAGTCAG CCTTACCAAGT AAAAAAACCT ATTAAAAAAC ACCACTCGAC
35401 ACGGCACCAAG CTCAATCACT CACAGTGTA AAAGGGCCAA GTACAGAGCG AGTATATATA
35461 GGACTAAAAA ATGACGTAAC GGTTAAAGTC CACAAAAACC ACCCAGAAAA CCGCACGCGA
35521 ACCTACGCC AGAAACGAAA GCCAAAAAC CCACAACTTC CTCAAATCTT CACTTCCGTT
35581 TTCCCACGAT ACGTCACTTC CCATTTAAA AAAAAACTAC AATTCCAAT ACATGCAAGT
35641 TACTCCGCC TAAAACCTAC GTCACCCGCC CCGTTCCAC GCCCCGCGCC ACGTCACAAA
35701 CTCCACCCCC TCATTATCAT ATTGGCTTCA ATCCAAAATA AGGTATATTA TTGATGATG

FIG. 70

1 CATCATCAAT AATATACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
 61 TTGTGACGTG CGCGGGGGCG TGGGAACGGG GCGGGTGACG TAGTAGTGTG CGGAAAGTGT
 121 GATGTTGCAA GTGTGGCGGA ACACATGTAA CGCACGGATG TGGCAAAAGT GACGTTTTG
 181 GTGTGCGCCG GTGTACACAG GAAGTGACAA TTTTCGCGCG GTTTTAGGCG GATGTTGTAG
 241 TAAATTGGGG CGTAACCGAG TAAGATTGCG CCATTTTGC GGGAAAATG AATAAGAGGA
 301 AGTGAATCT GAATAATTGT GTGTTACTCA TAGCGCGTAA TATTGTCGA GGGCCGCGGG
 361 GACTTGACCC GTTTACGTGG AGACTCGCCCG AGGTGTTTT CTCAGGTGT TTCCGCGTC
 421 CGGGTCAAAG TTGGCGTTT ATTATTATAG TCAGCTGACG TGTAGTGTAT TTATACCCGG
 481 TGAGTTCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTCTCC TCCGAGCCG
 541 TCCGACACCG GGACTGAAA TGAGACATAT TATCTGCCAC GGAGGTGTTA TTACCGAAGA
 601 AATGGCCGCC AGTCTTTGG ACCAGCTGAT CGAAGAGGT A CTGGCTGATA ATCTTCCACC
 661 TCCTAGCCAT TTTGAACCCAC CTACCCCTCA CGAACGTAT GATTAGACG TGACGGCCCC
 721 CGAAGATCCC AACGAGGAGG CGGTTTCGCA GATTTTCC GACTCTGTAA TGTTGGCGGT
 781 GCAGGAAGGG ATTGACTTAC TCACTTTCC GCGGGCGCC CGTCTCCCG AGCCGCTCA
 841 CCTTCCCAGC CAGCCCGAGC AGCCGGAGCA GAGAGCCTG GGTCCGGTTT CTATGCCAA
 901 CCTTGTACCG GAGGTGATCG ATCTTACCTG CACAGGGCT GGCTTCCAC CGAGTACGA
 961 CGAGGATGAA GAGGGTGAGG AGTTTGTGTT AGATTATGTG GAGCACCCCG GGCACGGTTG
 1021 CAGGCTTGTG CATTATCACC GGAGGAATAC GGGGGACCCA GATATTATGT GTTCGTTTG
 1081 CTATATGAGG ACCTGTGCA TGTTTGTCTA CAGTAAGTGA AAATTATGGG CAGTGGGTGA
 1141 TAGAGTGGTG GGTTTGGGTG CGTAATTGTT TTTTAATTG TTACAGTTT GTGGTTTAA
 1201 GAATTTGTA TTGTGATTGTT TTTAAAAGGT CCTGTGTCG AACCTGAGCC TGAGCCCGAG
 1261 CCAGAACCGG AGCCTGCAAG ACCTACCCGC CGTCCTAAAA TGGCCCTGC TATCTGAGA
 1321 CGCCCGACAT CACCTGTGTC TAGAGAATGC AATAGTAGTA CGGAATAGCTG TGACTCCGGT
 1381 CCTCTAAACA CACCTCCTGA GATACACCCCG GTGGTCCCGC TGTGCCCAT TAAACCAAGT
 1441 GCCGTGAGAG TTGGTGGCGC TCGCAGGCT GTGGAATGTA TCGAGGACTT GCTTAACGAG
 1501 CCTGGGCAAC CTTTGGACTT GAGCTGTAAA CGCCCCCAGGC CATAAGGTGT AAACCTGTGA
 1561 TTGCGTGTGT GGTTAACCCCG TTGTTTGCT GAATGAGTTG ATGTAAGTTT AATAAAGGGT
 1621 GAGATAATGT TTAACCTGCA TGGCGTGTAA AATGGGGCGG GGCTTAAAGG GTATATAATG
 1681 CGCCGTGGGC TAATCTTGGT TACATCTGAC CTCATGGAGG CTTGGAGTG TTGGAAGAT
 1741 TTTCTGCTG TGCGTAACAG GCTGAACAG AGCTCTAACAA GTACCTCTG GTTTTGGAGG
 1801 TTTCTGCTGG GCTCATCCCA GGAAAGTTA GTCTGCAGAA TPAAGGAGGA TACAAGTGG
 1861 GAATTTGAAG AGCTTTGAA ATCCGTGGT GAGCTGTTG ATTCTTGAA TCTGGTCAC
 1921 CAGGGCCTTT TCCAAGAGAA GGTCACTCAAG ACTTTGGATT TTTCCACACC GGGCGCGCT
 1981 GCGGCTGCTG TTGCTTTT GAGTTTATA AAGGATAAA GGAGCGAAG AACCCATCTG
 2041 AGCGGGGGGT ACCTGCTGGA TTTCTGGCC ATGCATCTGT GGAGAGCGGT TGTGAGACAC
 2101 AAGAATCGCC TGCTACTGTT GTCTCCGTC CGCCCGGGGA TAATACCGAC GGAGGAGCAG
 2161 CAGCAGCAGC AGGGAGAAC CAGGGGGCGG CGGCAGGAGC AGAGCCCATG GAACCCGAGA
 2221 GCCGGCCTGG ACCCTCGGA ATGAATGTT TACAGGTGGC TGAACGTAT CCAGAACTGA
 2281 GACGCATTTT GACAATTACA GAGGATGGGC AGGGGCTAAA GGGGGTAAAG AGGGAGCGGG
 2341 GGGCTTGTGA GGCTACAGAG GAGGCTAGGA ATCTAGCTT TAGCTTAATG ACCAGACACC
 2401 GTCTGAGTG TATTACTTT CAACAGATCA AGGATAATTG CGCTAATGAG CTTGATCTGC
 2461 TGGCCAGAA GTATTCCATA GAGGAGCTGA CCACCTACTG GCTGCAGCCA GGGGATGATT
 2521 TTGAGGAGGC TATTAGGTT TATGCAAAGG TGGCACTTAG GCCAGATTGC AAGTACAAGA
 2581 TCAGCAAATC TGTAATATC AGGAATTGTT GCTACATTTC TGGGAACGGG GCCGAGGTGG
 2641 AGATAGATAC GGAGGATAGG GTGGCCTTTA GATGTAGCAT GATAAAATATG TGGCCGGGG
 2701 TGCTTGGCAT GGACGGGGTG GTTATTATGA ATGTAAGGTT TACTGGCCCC AATTITAGCG
 2761 GTACGGTTT CCTGGCAAT ACCAACCTTA TCCTACACGG TGTAAGCTTC TATGGTTTA
 2821 ACAATACCTG TGTGGAAGCC TGGACCGATG TAAGGGTTCG GGGCTGTGCC TTTTACTGCT
 2881 GCTGGAAGGG GGTGGTGTGT CGCCCCAAAA GCAGGGCTTC AATTAAGAAA TGCCTCTTTG
 2941 AAAGGTGTAC CTTGGGTATC CTGCTGAGG GTAACCTCAG GGTGCGCCAC AATGTGGCCT
 3001 CCGACTGTGG TTGCTTCATG CTAGTGAAGG GCGTGGCTGT GATTAAGCAT AACATGGTAT
 3061 GTGGCAACTG CGAGGACAGG GCCTCTCAGA TGCTGACCTG CTCGACGCC AACTGTCACC
 3121 TGCTGAAGAC CATTACGTA GCCAGCCACT CTCGCAAGGC CTGGCCAGTG TTTGAGCATA
 3181 ACATACTGAC CCGCTGTCC TTGCAATTGG GTAACAGGAG GGGGGTGTTC CTACCTTAC
 3241 AATGCAATTG GAGTCACACT AAGATATTGCA TTGAGCCCGA GAGCATGTCC AAGGTGAACC

58/92

3301 TGAACGGGGT GTTTGACATG ACCATGAAGA TCTGGAAGGT GCTGAGGTAC GATGAGACCC
 3361 GCACCAAGGTG CAGACCCCTGC GAGTGTGGCG GTAAACATAT TAGGAACCAAG CCTGTGATGC
 3421 TGGATGTGAC CGAGGAGCTG AGGCCCCGATC ACTTGGTGCT GGCCTGCACC CGCGCTGAGT
 3481 TTGGCTCTAG CGATGAAGAT ACAGATTGAG GTACTGAAAT GTGTGGCGT GGCTTAAGGG
 3541 TGGGAAAGAA TATATAAGGT GGGGGCTTTA TGTAGTTTG TATCTGTTT GCAGCAGCC
 3601 CCGCCGCCAT GAGCACCAAC TCGTTGATG GAAGCATTGT GAGCTCATAT TTGACAACGC
 3661 GCATGCCCTT ATGGGCCGGG GTGCGTCAGA ATGTGATGGG CTCCAGCATT GATGGTCGCC
 3721 CCGTCCTGCC CGCAAACCTCT ACTACCTTGA CCTACGAGAC CGTGCTCTGG ACGCCGTTGG
 3781 AGACTGCAGC CTCCGCCGCC GCTTCAGCCG CTGCAGCCAC CGCCCGCGGG ATTGTGACTG
 3841 ACTTGCTTT CCTGAGCCCG CTTGCAAGCA GTGCAGCTTC CGCTTCATCC GCCCAGGATG
 3901 ACAAGTTGAC GGCTCTTTG GCACAAATTGG ATTCTTTGAC CCGGAAACTT AATGTCGTTT
 3961 CTCACCGAGCT GTTGGATCTG CGCCACAGG TTTCTGCCCT GAAGGCTTCC TCCCCCTCCCA
 4021 ATGCCGTTTA AAACATATAAT AAAAACCCAG ACTCTGTTT GATTGGATC AAGCAAGTGT
 4081 CTTGCTGTCT TTATTTAGGG GTTTTGCCTG CGCCGGTAGGC CGGGGACCCAG CGGTCTCGGT
 4141 CGTTGAGGGT CCTGTGTATT TTTCAGGA CGTGGTAAAG GTGACTCTGG ATGTTAGAT
 4201 ACATGGGCAT AAGCCCGTCT CTGGGGTGGG GGTAGCACCA CTGCAGAGCT TCATGCTGCC
 4261 GGGTGGTGT GTAGATGATC CAGTCGTAGC AGGAGCGCTG GGCCTGGTGC CTAAAAATGT
 4321 CTTCAGTAG CAAGCTGATT GCCAGGGGCA GGCCTTGGT GAAAGTGTGTT ACAAAGCGGT
 4381 TAAGCTGGGA TGGGTGCATA CGTGGGGATA TGAGATGCAT CTTGGACTGT ATTTTTAGGT
 4441 TGGCTATGTT CCCAGCCATA TCCCTCCGGG GATTCATGTT GTGCAGAAC ACCAGCACAG
 4501 TGTATCCGGT GCACTTGGGA AATTGTCAT GTAGCTTAGA AGGAATGCG TGGAAAGAACT
 4561 TGGAGACGCC CTTGTGACCT CCAAGATTTT CCATGCAATC GTCCATAATG ATGGCAATGG
 4621 GCCCACGGGC GGCGGCCCTGG GCGAAGATAT TTCTGGGATC ACTAACGTCA TAGTTGTGTT
 4681 CCAGGATGAG ATCGTCATAG GCCATTTTTA CAAAGCGGGC GCGGAGGGTG CCAGACTGCG
 4741 GTATAATGGT TCCATCCGGC CCAGGGGGT AGTTACCCCT ACAGATTTCG ATTTCCCACG
 4801 CTTTGAGTTC AGATGGGGG ATCATGTCAG CCGTGGGGC GATGAAGAAA ACGGTTTCCG
 4861 GGGTAGGGGA GATCAGCTGG GAAGAAAGCA GGTTCCTGAG CAGCTGCAGC TTACCGCAGC
 4921 CGGTGGGCCCG GTAAATCACA CCTATTACCG GGTGCAACTG GTAGTTAAGA GAGCTGCAGC
 4981 TGCGTCATC CCTGAGCAGG GGGGCCACTT CGTTAACGAT GTCCCTGACT CGCATGTTT
 5041 CCCTGACCAA ATCCGCCAGA AGGCAGCTGC CGCCCAAGCGA TAGCAGTTCT TGCAAGGAAG
 5101 CAAAGTTTT CAACGGTTTG AGACCGTCAG CCGTAGGCAT GCTTTGAGC GTTGTACCAA
 5161 GCAGTTCCAG GCGGTCCCAC AGCTCGGTCA CCTGCTCTAC GGCATCTCGA TCCAGCATAT
 5221 CTCCCTCGTT CGCGGGTTGG GGCAGCTTTC GCTGTACGGC AGTAGTCGGT GCTCGTCCAG
 5281 ACGGGCCAGG GTCACTGCTT TCCACGGGCG CAGGGTCCCTC GTCAGCGTAG TCTGGGTCA
 5341 GGTGAAGGGG TCCGCTCCGG GCTGCGCGCT GGCGAGGGTG CGCTTGAGGC TGGTCTGCT
 5401 GGTGCTGAAG CGCTGCCGGT CTTGCCCTG CGCGCTGGCC AGGTAGCATT TGACCATGGT
 5461 GTCATAGTCC AGCCCCCTCCG CGGCAGTGGCC CTTGGCGCC AGCTTGCCCT TGGAGGAGGC
 5521 GCCGCACGGAG GGGCAGTGCAGA GACTTTGAG GGCAGTGGC TTGGCGCGA GAAATACCGA
 5581 TTCCGGGGAG TAGGCATCCG CGCCGAGGC CCCGCAGACG GTCTCGCATT CCACGAGCCA
 5641 GGTGAGCTCT GGCGGTTCCG GTCAAAAC CAGGTTTCCC CCATGCTTT TGATGCGTTT
 5701 CTTACCTCTG GTTTCCATGA GCGGGTGTCC ACGCTCGGTG ACGAAAAGGC TGTCCGTGTC
 5761 CCCGTATACA GACTTGAGAG GCCTGTCCTC GAGCGGTGTT CCGCGGTCCCT CCTCGTATAG
 5821 AAACCTGGAC CACTCTGAGA CAAAGGCTCG CGTCCAGGCC AGCACGAAGG AGGCTAAGTG
 5881 GGAGGGGTAG CGGTGCTTGT CCACTAGGGG GTCCACTCGC TCCAGGGTGT GAAGACACAT
 5941 GTGCCCTCT TCAGGCATCAA GGAAGGTGAT TGGTTTGTAG GTGAGGAGCCA CGTGACCCGG
 6001 TGTTCCTGAA GGGGGCTAT AAAAGGGGT GGGGGCGCGT TCGTCCTCAC TCTCTTCCGC
 6061 ATCGCTGTCT GCGAGGGCCA GCTGTTGGGG TGAGTACTCC CTCTGAAAAG CGGGCATGAC
 6121 TTCTGCGCTA AGATTGTCAG TTTCCAAAAA CGAGGAGGAT TTGATATTCA CCTGGCCCGC
 6181 GGTGATGCCT TTGAGGGTGG CGGCATCCAT CTGGTCAGAA AAGACAATCT TTTTGTGTC
 6241 AACCTTGGTG GCAAAACGACC CGTAGAGGGC GTGGACAGC AACTTGGCGA TGGAGGCGAG
 6301 GGTGGTTT TTGTCGGAT CGGCAGCGCTC CTGGGGCGT ATGTTAGCT GCACGTATTC
 6361 GCGCGCAACG CACCGCCATT CGGGAAAGAC GGTGGTGCAGC TCGTCGGGCA CCAGGTGCAC
 6421 GCGCCAACCG CGGTTGTGCA GGGTGACAAG GTCAACGCTG GTGGCTACCT CTCCGGTAG
 6481 GCGCTCGTTG GTCCAGCAGA GGCGGCCGCC CTTGCGCGAG CAGAATGGCG GTAGGGGGTC
 6541 TAGCTGCGTC TCGTCCGGGG CGTCTGCGTC CACGGTAAAG ACCCCGGGCA GCAGGCGCGC

FIG. 8B

59/92

6601 GTCGAAGTAG TCTATCTTGC ATCCTTCAA GTCTAGCGCC TGCTGCCATG CGCGGGCGGC
 6661 AAGCGCGCGC TCGTATGGGT TGAGTGGGGG ACCCCATGGC ATGGGGTGGG TGAGCGCGGA
 6721 GCGTACATG CCGAAATGT CGTAAACGTA GAGGGGCTCT CTGAGTATTG CAAGATATGT
 6781 AGGGTAGCAT CTTCCACCGC GGATGCTGCC GCGCACGTA TCGTATAGTT CGTGCAGGG
 6841 AGCGAGGAGG TCGGGACCGA GGTTGCTACG GCGGGGCTGC TCTGCTCGGA AGACTATCTG
 6901 CCTGAAGATG GCATGTAGT TGGATGATAT GTTGGACGC TGGAAAGACGT TGAAGCTGGC
 6961 GTCTGTGAGA CCTACCGCGT CACCGACGAA GGAGGCGTAG GAGTCGCAGA GCTTGTGAC
 7021 CAGCTCGCG GTGACCTGCA CGTCTAGGGC GCAGTAGTCC AGGGTTTCCCT TGATGATGTC
 7081 ATACTTATCC TGTCCCTTT TTTTCCACAG CTGCGGGTTG AGGACAAACT CTTCGCGGTC
 7141 TTTCACTAC TCTTGGATCG GAAACCCGTC GGCCTCCGAA CGGTAAGAGC CTAGCATGTA
 7201 GAACTGGTTG ACGGCCTGGT AGGCGCAGCA TCCCTTTCT ACGGGTAGCG CGTATGCCG
 7261 CGCGGCCCTTC CGGAGCGAGG TGTGGGTGAG CGCAAAGGTG TCCCTGACCA TGACTTTGAG
 7321 GTACTGGTAT TTGAAGTCAG TGTGTCGCA TCCGCCCTGC TCCCAGAGCA AAAAGTCCGT
 7381 GCGCTTTTG GAACGCGGAT TTGGCAGGGC GAAGGTGACA TCGTTGAAGA GTATCTTCC
 7441 CGCGCGAGGC ATAAAGTTG GTGTGATGCG GAAGGGTCCC GGCACCTCGG AACGGTTGTT
 7501 AATTACCTGG CGGGCGAGCA CGATCTCGTAAAGCCGTG ATGTTGTGGC CCACAATGTA
 7561 AAGTCCAAG AAGCGCGGGA TGCCCTTGAT GGAAGGCAAT TTTTAAGTT CCTCGTAGGT
 7621 GAGCTCTTCAGGGGAGCTGA GCGGCCCTGC TGAAAGGGGCA CAGTCTGCAA GATGAGGGTT
 7681 GGAAGCGACG AATGAGCTC ACAGGTACAG GGCCATTAGC ATTTGCAAGGT GGTGCGAAA
 7741 GGTCTAACCTGGTGGCAGCTA TGGCATTTC TTCTGGGTG ATGCACTAGA AGGTAAGCGG
 7801 GTCTGTTC CAGCGGTCCC ATCCAAGGTT CGCGGCTAGG TCTCGCGCGG CAGTCACTAG
 7861 AGGCTCATCT CCGCGAAGCT TCATGACCAAG CATGAAGGGC ACGAGCTGCT TCCCAAAGGC
 7921 CCCCATCCAA GTATAGGTCT CTACATCGTA GGTGACAAG AGACGCTCGG TGCGAGGATG
 7981 CGAGCCGATC GGGAAAGAACT GGATCTCCCG CCACCAATTG GAGGAGTGGC TATTGATGTC
 8041 GTGAAAGTAG AAGTCCCTGC GACGGGCCGA ACACTCGTC TGGCTTTTGT AAAACGTGC
 8101 GCAGTACTGG CAGCGGTGCA CGGGCTGTAC ATCCTGCAAG AGGTTGACCT GACGACCGCG
 8161 CACAAGGAAG CAGAGTGGGA ATTTGAGCCC CTGCGCTGGC GGGTTTGGCT GGTGGTCTTC
 8221 TACTTCGGCT GCTTGTCCCT GACCGTCTGG CTGCTCGAGG GGAGTTACGG TGGATCGGAC
 8281 CACCAACGCCG CGCGAGCCCA AAGTCCAGAT GTCCCGCGC GGGCGTCGGG CTTGATGAC
 8341 AACATCGCGC AGATGGGAGC TGTCATGGT CTGGAGCTCC CGCGCGTCA GGTCAAGCGG
 8401 GAGCTCTGC AGGTTTACCT CGCATAGACG GGTGAGGGC CGGGCTAGAT CCAGGTGATA
 8461 CCTAATTCTC AGGGGCTGGT TGGTGGCGGC GTCGATGGCT TGCAAGAGGC CGCATCCCCG
 8521 CGCGCGCACT ACGGTACCGC GCGGGGGCG GTGGGCCCG GGGGTGTCCT TGGATGATGC
 8581 ATCTAAAGC GGTGACCGGG CGAGGCCCCC GGAGGTAGGG GGGGCTCCGG ACCCGCCGGG
 8641 AGAGGGGGCA GGGGCACGTC GGCAGCGC GCGGGCAGGA GCTGGTGCTG CGCGCGTAGG
 8701 TTGCTGGCGA ACAGCAGCAC GCGGGGGTTG ATCTCCTGAA TCTGGCCCT CTGCGTGAAG
 8761 ACGACGGGCC CGGTGAGCTT GAGCTGAAA GAGAGTTCGA CAGAATCAAT TTGGTGTGCG
 8821 TTGACGGCGG CCTGGCGAA AATCTCCTGC ACGTCTCTG AGTTGTCTTG ATAGGCGATC
 8881 TCGGCCATGA ACTGCTCGAT CTCTCTCTC TGGAGATCTC CGCGTCCGGC TCGCTCCACG
 8941 GTGGCGCGA GGTGTTGGA AATGCGGGCC ATGAGCTGCG AGAACGGCGTT GAGGGCTCCC
 9001 TCGTTCAGA CGCGGCTGTA GACCAACGCC CTTTCCGGCAT CGCGGGCGCG CATGACCAAC
 9061 TGCGCGAGAT TGAGCTCAC GTGCCGGCG AAGACGGCGT AGTTTCCGAG CGCGTGAAG
 9121 AGGTAGTTGA GGGTGTGGC GGTGTGTTCT GCCACGAAGA AGTACATAAC CCAGCGTCGC
 9181 AACGTGGATT CGTTGATATC CCCCAGGCC TCAAGGGCGCT CCATGGCCTC GTAGAAGTCC
 9241 AGCGCGAAGT TGAAAAGCTG GGAGTTGCGC GCGGACACGG TTAACCTCTC CTCCAGAAGA
 9301 CGGATGAGCT CGGCGACAGT GTCGCGCACC TCGCGCTCAA AGGCTACAGG GGCCTCTCT
 9361 TCTTCTTCAA TCTCCTCTC CATAAGGGCC TCCCCCTCTT CTTCTCTGG CGGGCGTGGG
 9421 GGAGGGGGGA CACGGCGCG ACGACGGCG ACCGGGAGGC GGTGACACAA GCGCTCGATC
 9481 ATCTCCCCGC GCGGACGGCG CATGGCTCG GTGACGGCGC GGCGCTCTC GCGGGGGCGC
 9541 AGTTGGAAGA CGCCGCCCCGT CATGCCCCGG TTATGGGTG GCGGGGGCGT GCCATGCGGC
 9601 AGGGATACGG CGCTAACGAT GCATCTAAC AATTGTTGTG TAGGTACTCC CGCGCGAGG
 9661 GACCTGAGCG AGTCCGATC GACCGGATCG GAAAACCTCT CGAGAAAGGC GTCTAACCAAG
 9721 TCACAGTCGC AAGGTAGGCT GAGCACCGTG GCGGGCGGCA CGGGCGGGCG GTCGGGGTTG
 9781 TTCTGGCGG AGGTGCTGCT GATGATGTAA TTAAAGTAGG CGGTCTTGAG ACGGGGGATG
 9841 GTCGACAGAA GCACCATGTC CTTGGGTCCG GCCTGCTGAA TGCGCAGGCG GTCGGCCATG

FIG. 8C

60/92

9901	CCCCAGGCTT	CGTTTTGACA	TCGGCGCAGG	TCTTTGTAGT	AGTCTGCAT	GAGCCTTCT
9961	ACCGGCACTT	CTTCTCTCTCC	TTCCCTCTGT	CCTGCATCTC	TTGCATCTAT	CGCTCGGGCG
10021	GCGGGGGAGT	TTGGCCGTAG	GTGGCGCCCT	CTTCCTCCCA	TCCGTGTGAC	CCCGAAGCCC
10081	CTCATCGGCT	GAAGCAGGGC	TAGTCGGCG	ACAACCGCGT	CGGCTAATAT	GGCCTGCTG
10141	ACCTGCGTGA	GGGTAGACTG	GAAGTCATCC	ATGTCACAAA	AGCGGTGGTA	TGGCCCGCGT
10201	TTGATGGGT	AAGTGCAGTT	GGCCATAACG	GACCAATTAA	CGGTCTGGT	ACCCGGCTG
10261	GAGAGCTCGG	TGTACCTGAG	ACCGCAGTAA	GCCCTCGAGT	CAAATACGTA	GTCTGTTGCAA
10321	GTCCGCACCA	GGTACTGGTA	TCCCACCAAA	AAAGTGCAGGC	GCGGCTGGCG	GTAGAGGGC
10381	CAGCGTAGGG	TGGCCGGGGC	TCCGGGGCG	AGATCTTCCA	ACATAAGGC	ATGATATCCG
10441	TAGATGTACC	TGGACATCCA	GGTGATGCCG	GCGGCGGTGG	TGGAGGCGCG	CGGAAAGTCG
10501	CGGACGCGGT	TCCAGATTT	GCGCAGCGGC	AAAAAGTGT	CCATGGTCGG	GACGCTCTGG
10561	CCGGTCAGGC	GCGCAGAAC	GTGACGCTC	TAGACCGTGC	AAAAGGAGAG	CCTGTAAGCG
10621	GGCACTCTTC	CGTGGTCTGG	TGGATAAAATT	CGCAAGGGTA	TCATGGCGGA	CGACCGGGGT
10681	TCGAGCCCCG	TATCCGGCCG	TCCCGCGTGA	TCCATGCCGT	TACCGCCCGC	GTGTCGAACC
10741	CAGGTGTGCG	ACGTCAGACA	ACGGGGGAGT	GCTCCTTTG	GCTTCCTTCC	AGGCGCGGGC
10801	GCTGCTGCG	TAGCTTTTT	GGCCACTGGC	GCGCCCGCAGC	GTAAGGGTT	AGGCTGGAAA
10861	GGCAAAGCAT	TAAGTGGCTC	GCTCCCTGTA	GCGGGAGGGT	TATTTTCCAA	GGGTTGAGTC
10921	GGGGGACCCC	CGGTTCGAGT	CTCGGACCGG	CGGGACTGCG	GCGAACGGGG	GTGTCCTCC
10981	CCGTCATGCA	AGACCCCCCT	TGCAAATTCC	TCCGGAAAC	GGGACGAGCC	CTTTTTTTGC
11041	TTTTCCCAGA	TGCATCCGGT	GCTGGGGCAG	ATGCGCCCCC	CTCCTCAGCA	GCGGCAAGAG
11101	CAAGAGCAGC	GGCAGACATG	CAGGGCACCC	TCCCCCTCCTC	CTACCGCGTC	AGGAGGGGCC
11161	ACATCCGCGG	TTGACGCGGC	AGCAGATGGT	GATTACGAAC	CCCCCGGGCG	CCGGGGCCCCG
11221	CACTACCTGG	ACTTGGAGGA	GGGCGAGGGC	CTGGCGCGGC	TAGGAGCGCC	CTCTCCTGAG
11281	CGGTACCCAA	GGGTGCAAGCT	GAAGCGTGAT	ACGCGTGAGG	CGTACCGTGC	GCGGCAGAAC
11341	CTGTTTCGCG	ACCGCGAGGG	AGAGGAGGCC	GAGGAGATGC	GGGATCGAAA	GTTCACCGCA
11401	GGGCGCGAGC	TGCGGCATGG	CCTGAATCGC	GAGCGGTTGC	TGCGCGAGGA	GGACTTTGAG
11461	CCCGACGCGC	GAACCGGGAT	TAGTCCCGCG	CGCGCACACG	TGGCGCCCGC	CGACCTGGTA
11521	ACCGCATACG	ACGACAGCGT	GAACCGAGAG	ATTAACATT	AAAAAGCTT	TAACAACCCAC
11581	GTGCGTACGC	TTGTGGCGCG	CGAGGAGGTG	GCTATAGGAC	TGATGCATCT	GTGGGACTTT
11641	GTAAGCGCGC	TGGAGCAAAA	CCCAAATAGC	AAGCCGCTCA	TGGCGCAGCT	GTTCCTTATA
11701	GTGCAAGCACA	GCAGGGACAA	CGAGGCATTC	AGGGATGCCG	TGCTAAACAT	AGTAGAGCCC
11761	GAGGGCCGCT	GGCTGCTCGA	TTTGATAAAAC	ATCCCTGCAGA	GCATAGTGGT	GCAGGAGGCC
11821	AGCTTGAGCC	TGGCTGACAA	GGTGGCCGCC	ATCAACTATT	CCATGCTTAG	CCTGGGCAAG
11881	TTTACGCC	GCAAGATATA	CCATACCCCT	TACGTTCCCA	TAGACAAGGA	GGTAAAGATC
11941	GAGGGGTTCT	ACATGCCAT	GGCGCTGAAG	GTGCTTACCT	TGAGCGACGA	CCTGGGCGTT
12001	TATCGAACG	AGCGCATCCA	CAAGGCCGTG	AGCGTGAGCC	GGCGGGCGGA	GCTCAGCGAC
12061	CGCGAGCTGA	TGCACAGCCT	GCAAAGGGC	CTGGCTGGCA	CGGGCAGCGG	CGATAGAGAG
12121	GCCGAGTCCT	ACTTTGACGC	GGGGCGCTGAC	CTGGCGCTGGG	CCCCAAGCCC	ACGGCGCCCTG
12181	GAGGCAGCTG	GGGGCGGCC	TGGGCTGGCG	GTGGCACCCC	CGCGCGCTGG	CAACGTCGGC
12241	GGCGTGGAGG	ATATGACGA	GGACGATGAG	TACGAGCCAG	AGGACGGCGA	GTACTAAGCG
12301	GTGATGTTTC	TGATCAGATG	ATGCAAGACG	CAACGGACCC	GGCGGTGCGG	GCGGCGCTGC
12361	AGAGCCAGCC	GTCCGGCCCT	AACTCCACGG	ACGACTGGCG	CCAGGTCA	GACCGCATTCA
12421	TGTCGCTGAC	TGCGCGCAAT	CCTGACCGT	TCCGGCAGCA	CCCGCAGGCC	AACCGGCTCT
12481	CCGCAATTCT	GGAAAGCGGT	GTCCCGGCC	GGCAGAACCC	CACGCACGAG	AAGGTGCTGG
12541	CGATCGTAAA	CGCGCTGGCC	GAAAACAGGG	CCATCCGGCC	CGACGAGGCC	GGCCTGGTCT
12601	ACGACGCGCT	GCTTCAGCGC	GTGGCTCGTT	ACAACAGCGG	CAACGTGCAG	ACCAACCTGG
12661	ACCGGCTGGT	GGGGGATGTG	CGCGAGGCCG	TGGCGCAGCG	TGAGCGCGCG	CAGCAGCAGG
12721	GCAACCTGGG	CTCCATGGTT	GCACATAACG	CCTTCCTGAG	TACACAGCCC	GCCAACGTGC
12781	CGCGGGGACA	GGAGGACTAC	ACCAACTTG	TGAGCGCACT	GGGGCTAATG	GTGACTGAGA
12841	CACCGCAAAG	TGAGGTGTAC	CAGTCCTGGG	CAGACTATT	TTTCCAGACC	AGTAGACAAAG
12901	GCCTGCAAGAC	CGTAAACCTG	AGCCAGGCTT	TCAAAAACCTT	CCAGGGCTG	TGGGGGGGTG
12961	GGGCTCCAC	AGGCACCGC	GCGACCGTGT	CTAGCTTGCT	GACGCCAAC	TGGCGCCTGT
13021	TGCTGCTGCT	AATAGCGCCC	TTCACGGACA	GTGGCAGCGT	GTCCCCGGAC	ACATACCTAG
13081	GTCACTTGCT	GACACTGTAC	CGCGAGGCCA	TAGGTCAGGC	GCATGTGGAC	GAGCATACTT
13141	TCCAGGAGAT	TACAAGTGT	AGCCCGCGC	TGGGGCAGGA	GGACACGGGC	AGCCTGGAGG

FIG. 8D

13201 CAACCCCTAAA CTACCTGCTG ACCAACCGGC GGCAGAAAGAT CCCCTCGTTG CACAGTTAA
 13261 ACAGCGAGGA GGAGCGCATT TTGCGCTACG TGCAGCAGAG CGTGAGCCCTT AACCTGATGC
 13321 GCGACGGGGT AACGCCAGC GTGGCGCTGG ACATGACCCGC GCGAACATG GAACCGGGCA
 13381 TGTATGCCTC AAACCGGCCG TTTATCAACC GCCTAATGGA CTACTTCAT CGCGCGGCC
 13441 CCGTGAACCC CGAGTATTC ACCAATGCCA TCTTGAAACCC GCACTGGCTA CGGCCCCCTG
 13501 GTTTCTACAC CGGGGGATTC GAGGTGCCG AGGGTAACCGA TGGATTCCCTC TGGGACGACA
 13561 TAGACGACAG CGTGTTCCTC CCGAACCGC AGACCCCTGCT AGAGTTGCAA CAGCGCGAGC
 13621 AGGCAGAGGC GGCGCTGCCA AAGGAAAGCT TCCGCAGGCC AAGCAGCTTG TCCGATCTAG
 13681 GCGCTGCCG CCCGCGGTCA GATGCTAGTA GCCCCATTTC AAGCTTGATA GGGTCTCTTA
 13741 CCAGCACTCG CACCACCCGC CGCGCCCTGC TGGGCGAGGA GGAGTACCTA ACAACTCCG
 13801 TGCTGCAGCC GCAGCGCGAA AAAAACCTGC CTCCGGCATT TCCCAACAAAC GGGATAGAGA
 13861 GCCTAGTGGG CAAGATGAGT AGATGGAAGA CGTACGCCA GGAGCACAGG GACGTGCCAG
 13921 GCCCGCGCCG GCCCACCCGT CGTCAAAGGC AGCACCGTCC TGGATTGAGG AAGGAGTGGC AAGCCGTTG
 13981 ACGATGACTC GGCAGACGAC AGCACCGTCC TGGATTGAGG AAGGAGTGGC AAGCCGTTG
 14041 CGCACCTTCG CCCCAGCGT GGGAGAATGT TTTAAAAAAA AAAAGCATG ATGCAAAATA
 14101 AAAAACCTCAC CAAGGCCATG GCACCCAGCG TTGGTTTCT TGTATTCCCC TTAGTATGCG
 14161 GCGCGCGGGCG ATGTATGAGG AAGGTCCCTC TCCCTCCCTAC GAGAGTGTGG TGAGCGCGGC
 14221 GCCAGTGGCG GCGGCGCTGG GTTCTCCCTT CGATGCTCCC CTGGACCCGC CGTTTGTGCC
 14281 TCCCGGGTAC CTGGCGCTA CGGGGGGGAG AAACAGCATC CGTTACTCTG AGTTGGCACC
 14341 CCTATTGAC ACCACCCGTG TGTACCTGGT GGACAACAAG TCAACGGATG TGGCATCCCT
 14401 GAACTACCAG AACGACCACA GCAACTTTCT GACCACGGTC ATTCAAAACA ATGACTACAG
 14461 CCCGGGGGAG GCAAGCACAC AGACCATCAA TCTTGACGAC CGGTCGCACT GGGCGGGCGA
 14521 CCTGAAAACC ATCCTGCATA CCAACATGCC AAATGTGAAC GAGTCATGT TTACCAATAA
 14581 GTTTAAGGCG CGGGTGATGG TGTGGCGCTT GCCTACTAAG GACAATCAGG TGGAGCTGAA
 14641 ATACGAGTGG GTGGAGTTCA CGCTGCCCGA GGGCAACTAC TCCGAGACCA TGACCATAGA
 14701 CCTTATGAAAC AACGCGATCG TGGAGCACTA CTTGAAAGTG GGCAGACAGA ACGGGGTTCT
 14761 GGAAAGCGAC ATCGGGGTTAA AGTTTGACAC CGCGCAACTC AGACTGGGT TTGACCCCCG
 14821 CACTGGCTT GTCATGCCCTG GGGTATATAC AAACGAAGCC TTCCATCCAG ACATCATTT
 14881 GCTGCCAGGA TGCGGGGGTGG ACTTCACCCA CAGCCGCCCTG AGCAACTTGT TGGGCATCCG
 14941 CAAGCGGAA CCCTTCCAGG AGGGCTTCTAG GATCACCTAC GATGATCTGG AGGGTGGTAA
 15001 CATTCCCGCA CTGTTGGATG TGGACGCCA CCAGGCAGGC TTGAAAGATG ACACCGAACAA
 15061 GGGCGGGGGT GGCGCAGGGC GCAGAACAG CAGTGGCAGC GGCGCGGAAG AGAACTCCAA
 15121 CGCGGCAGCC GCGGCAATGC AGCCCGTGGA GGACATGAAC GATCATGCCA TTGCGGGCGA
 15181 CACCTTGCC ACACGGGCTG AGGAGAAGCG CGCTGAGGCC GAAGCAGCGG CCGAAGCTGC
 15241 CGCCCCCGCT GCGCAACCCG AGGTCGAGAA GCCTCAGAAAG AAACCGGTGA TCAAAACCCCT
 15301 GACAGAGGAC AGCAAGAAC GCAGTTACAA CCTAATAAGC AATGACAGCA CCTTCACCCA
 15361 GTACCGCAGC TGGTACCTTG CATACAACTA CGCGCACCCCT CAGACCGGAA TCCGCTCATG
 15421 GACCTGCTT TGCACTCCCTG ACGTAACCTG CGGCTCGGAG CAGGCTACT GGTGTTGCC
 15481 AGACATGATG CAAGACCCCG TGACCTCCCG CTCCACGCCA CAGATCAGCA ACTTTCCGGT
 15541 GGTGGCGGCC GAGCTGTTGC CGTGCACTC CAAGAGCTC TACAACGACC AGGCCGTCTA
 15601 CTCCCAACTC ATCCGCCAGT TTACCTCTCT GACCCACGTG TTCAATCGCT TTCCCGAGAA
 15661 CCAGATTTG GCGGCCCGC CAGCCCCCAC CATCACCAAC GTCAGTAAA ACGTTCCCTGC
 15721 TCTCACAGAT CACGGGACGC TACCGCTGCG CAACAGCATC GGAGGAGTCC AGCGAGTGA
 15781 CATTACTGAC GCCAGACGCC GCACCTGCCCT ACAGTGTTC AAGGGCCCTGG GCATAGTCTC
 15841 GCCCGCGTC CTATCGAGCC GCACCTTTG AGCAAGCATG TCCATCCTTA TATCGCCCAAG
 15901 CAATAACACA GGCTGGGGCC TGCGCTTCCC AAGCAAGATG TTTGGCGGGG CCAAGAAGCG
 15961 CTCCGACCAA CACCCAGTGC CGTGCGCGG GCACTACCGC GCGCCCTGGG GCGCGCACAA
 16021 ACGCGGCCGC ACTGGGCGCA CCACCGTCGA TGACGCCATC GACGCCGTGG TGGAGGAGGC
 16081 GCGCAACTAC ACGCCACCGC CGCCACCAAGT GTCCACAGTG GACGCCGCCA TTCAAGACCGT
 16141 GGTGCGCGGA GCGCGGCCGT ATGCTAAAAT GAAGAGACGG CGGAGGCGCG TACCGCTCG
 16201 CCACCGCCGC CGACCCGGCA CTGCGGCCCA ACGGCGGGCG GCGGCCCTGC TTAACCGGCC
 16261 ACGTCGCAACCGC GGCGGACGGG CGGCCATGCG GGCGCGCTGA AGGCTGGCCG CGGGTATTGT
 16321 CACTGTGCCCTC CCCAGGTCCA GGCGACGAGC GGCGGCCGCC GCAAGCGCGG CCAATTAGTGC
 16381 TATGACTCAG GGTGCGAGGG GCAACGTGTA TTGGGTGCGC GACTCGGTTA GCGGGCCTGCG
 16441 CGTGCCCGTG CGCACCCGCC CCCCCGCGAA CTAGATTGCA AGAAAAAAACT ACTTAGACTC

62/92

16501 GTACTGTTGT ATGTATCCAG CGGCAGCGGC GCGCAACGAA GCTATGTCGA AGCGCAAAAT
 16561 CAAAGAAGAG ATGCTCCAGG TCATCGCGCC GGAGATCTAT GGCCCCCGA AGAAGGAAGA
 16621 GCAGGATTAC AAGCCCCGAA AGCTAAAGCG GGTCAAAAAG AAAAAGAAAG ATGATGATGA
 16681 TGAACTTGAC GACGAGGTGG AACTGCTGCA CGCTACCGCG CCCAGGCGAC GGGTACAGTG
 16741 GAAAGGTCGA CGCGTAAAC GTGTTTGC ACCCGGCACC ACCGTAGTCT TTACGCCCCG
 16801 TGAGCGCTCC ACCCGCACCT ACAAGCGCGT GTATGATGAG GTGTACGGCG ACCAGGACCT
 16861 GCTTGAGCAG GCCAACGAGC GCCTCGGGGA GTTTGCCTAC GGAAAGCGGC ATAAGGACAT
 16921 GCTGGCGTTG CCGCTGGACG AGGGCAACCC AACACCTAGC CTAAAGCCCG TAACACTGCA
 16981 GCAGGTGCTG CCCCGCCTTG CACCGTCCGA AGAAAAGCGC GGCCTAAAGC GCGAGTCTGG
 17041 TGACTTGGCA CCCACCGTGC AGCTGATGGT ACCCAAGCGC CAGCGACTGG AAGATGTCTT
 17101 GGAAAAAATG ACCGTGGAAC CTGGGCTGGA GCCCGAGGTC CGCGTGCAGC CAATCAAGCA
 17161 GGTGGCGCCG GGACTGGCG TGCAAGACCGT GGACGTTCAAG ATACCCACTA CCAGTAGCAC
 17221 CAGTATTGCC ACCGCCACAG AGGGCATGGA GACACAAACG TCCCCGGTTG CCTCAGCGGT
 17281 GGCGGATGCC GCGGTGCGAGG CGGTCGCTGC GGCGCGCTCC AAGACCTCTA CGGAGGTGCA
 17341 AACGGACCCG TGGATGTTTC CGCTTCAAGC CCCCCGGCGC CGCGCGGTT CGAGGAAGTA
 17401 CGGCGCCGCC ACCGCGCTAC TGCCCCATA TGCCCTACAT CCTTCATTG CGCCCTACCC
 17461 CGGCTATCGT GGCTACACCT ACCGGCCCCAG AAGACGAGCA ACTACCCGAC GCGGAACCC
 17521 CACTGGAACCC CGCCGCCGCC GTCGCCGTGC CCAGCCCGTG CTGGCCCCGA TTTCGGTGC
 17581 CAGGGTGGCT CGCGAAGGAG GCAGGACCCCT GGTGCTGCCA ACAGCGCGCT ACCACCC
 17641 CATCGTTAA AAGCCGGTCT TTGTTGTTCT TGCAAGATATG GCCCTCACCT GCGCCCTCC
 17701 TTTCGGTGT CGGGGATTCC GAGGAAGAAT GCACCGTAGG AGGGGCATGG CGGGCACCG
 17761 CCTGACGGGC GGCATCGCTC GTGCGCACCA CGGGCGGCC CGCGCGTCGC ACCGTCGCAT
 17821 GCGCGGCGGT ATCCTGCCCT TCCTTATTCC ACTGATGCC GCGGGGATTG GCGCGGTG
 17881 CGGAATTGCA TCCGTGGCT TGCAGGCGCA GAGACACTGA TTAAAACAA GTTGCATG
 17941 GAAAATCAA ATAATAAAAGT CTGGACTCTC ACGCTCGCTT GGTCTCTGAA CTATTTGTA
 18001 GAATGGAAGA CATCAACTT GCGCTCTCTGG CCCCGCAGCA CGGCTCGCGC CGCTTCATGG
 18061 GAAACTGGCA AGATATCGGC ACCAGCAATA TGAGCGTTGG CGCCTTCAGC TGGGGCTCG
 18121 TGTTGAGCGG CATTAAAAAT TTGCGTTCCA CCGTTAAGAA CTATGGCAGC AAGGCTGGA
 18181 ACAGCAGCAC AGGCCAGATG CTGAGGGATA AGTTGAAAGA GCAAATTTT CAACAAAAGG
 18241 TGGTAGATGG CCTGGCTCT GGCATTAGCG GGGTGGTGA CCTGGCCAAC CAGGCAGTGC
 18301 AAAATAAGAT TAACAGTAAG CTTGATCCCC GCCCTCCCGT AGAGGAGCCT CCACCGGCC
 18361 TGGAGACAGT GTCTCCAGAG GGGCGTGGCG AAAAGCGTCC GCGCCCGAC AGGGAAAGAA
 18421 CTCTGGTGCAC GCAAATAGAC GAGCTCCCT CGTACGAGGA GGCACTAAAG CAAGGCTGC
 18481 CCACCCACCG TCCCACCGC CCCATGGCTA CGGGAGTGC GGGCAGCAC ACACCCGTA
 18541 CGCTGGACCT GCCTCCCCCG GCGACACCC AGCAGAAACC TGTGCTGCCA GCGCGACCG
 18601 CGCTGGTGTG AACCCGCTCT AGCCCGCGT CCCTGCGCC CGCCGCCAGC GGTCCGCGAT
 18661 CGTGTGGCC CGTAGCCAGT GGCAACTGGC AAGACACACT GAACAGCATC TGGGGTCTGG
 18721 GGGTGCAATC CCTGAAGCGC CGACGATGCT TCTGAATAGC TAACGTGTCG TATGTG
 18781 ATGTATGCGT CCATGTCGCC GCCAGAGGAG CTGCTGAGCC GCGCGCGCC CGCTTTCAA
 18841 GATGGCTACC CCTTCGATGA TGCCGAGTG GTCTTACATG CACATCTCGG GCCAGGACGC
 18901 CTGGAGTAC CTGAGCCCCG GGCTGGTGA GTTTGCCCG GGCACCGAGA CGTACTTCAG
 18961 CCTGAATAAC AAGTTAGAA ACCCCACGGT GGCGCTACG CACGACGTGA CCACAGACCG
 19021 GTCCCAGCGT TTGACGCTGC GGTTCATCCC TGTGGACCGT GAGGATACTG CGTACTCGTA
 19081 CAAGGCGCGG TTGACCCCTAG CTGTTGGTGA TAACCGTGTG CTGGACATGG CTTCCACGTA
 19141 CTTGACATC CGCGCGTGC TGGACAGGGG CCCTACTTT AAGCCCTACT CTGGCACTGC
 19201 CTACAACGCC CTGGCTCCCA AGGGTGGCCC AAATCCTGCA GAATGGGATG AACCTGCTAC
 19261 TGCTCTTGAA ATAAACCTAG AAGAAGAGGA CGATGACAAC GAAGACGAAG TAGACGAGCA
 19321 AGCTGAGCAG CAAAAAAACTC ACGTATTGG CGAGGCGCTT TATTCTGGTA TAAATATTAC
 19381 AAAGGAGGGT ATTCAAAATAG GTGTCGAAGG TCAAACACTT AAATATGCCG ATAAAACATT
 19441 TCAACCTGAA CCTCAAATAG GAGAATCTCA GTGGTACGAA ACTGAAATT AATCATGCAGC
 19501 TGGGAGAGTC CTTAAAAAGA CTACCCCAAT GAAACCATGT TACGGTTCAT ATGAAAACC
 19561 CACAAATGAA AATGGAGGGC AAGGCATTCT TGTAAAGCAA CAAAATGGAA AGCTAGAAG
 19621 TCAAGTGGAA ATGCAATTTC TCTCAACTAC TGAGGCGACC GCAGGCAATG GTGATAACTT
 19681 GACTCCTAAA GTGGTATTGT ACAGTGAAGA TGTAGATATA GAAACCCAG ACACTCATAT
 19741 TTCTTACATG CCCACTATTA AGGAAGGTA CTCACGAGAA CTAATGGGCC AACAACTAT

FIG. 8F

63/92

19801 GCCCAACAGG CCTAATTACA TTGCTTTAG GGACAATTT ATTGGTCTAA TGTATTACAA
 19861 CAGCACGGGT AATATGGGTG TTCTGGCGGG CCAAGCATCG CAGTTGAATG CTGTTGAGA
 19921 TTTGCAAGAC AGAAAACACAG AGCTTTCATA CCAGCTTTG CTTGATTCCA TTGGTGTAG
 19981 AACCAAGGTAC TTTCTATGT GGAATCAGGC TGTTGACAGC TATGATCCAG ATGTTAGAAT
 20041 TATTGAAAAT CATGGAACGT AAGATGAAC TCCAAATTAC TGCTTCCAC TGGGAGGTGT
 20101 GATTAATACA GAGACTCTA CCAAGGTAAA ACCTAAAACA GGTCAGGAAA ATGGATGGGA
 20161 AAAAGATGCT ACAGAATTTC CAGATAAAA TGAAATAAGA GTTGGAAATA ATTTTGCCAT
 20221 GGAATCAAT CTAATGCAC ACCTGTGGAG AAATTCCTG TACTCCAACA TAGCGCTGTA
 20281 TTTGCCGAC AAGCTAAAGT ACAGTCTTC CAACGTAAA ATTTCTGATA ACCCAAACAC
 20341 CTACGACTAC ATGAACAAGC GAGTGGTGGC TCCCAGGGTTA GTGGACTGCT ACATTAACCT
 20401 TGAGGACACGC TGGTCCCTG ACTATATGGA CAACTCAAC CCATTAAACC ACCACCGCAA
 20461 TGCTGGCCTG CGCTACCGCT CAAATGTTGCT GGCAATGGT CGCTATGTGC CCTTCCACAT
 20521 CCAGGTGCCT CAGAAGTTCT TTGCAATTAA AAACCTCCTT CTCCCTGGG GCTCATACAC
 20581 CTACGAGTGG AACTTCAGGA AGGATGTTAA CATGGTTCTG CAGAGCTCCC TAGGAAATGA
 20641 CCTAAGGGTT GACGGAGCCA GCATTAAGTT TGATAGCATT TGCCCTTACG CCACCTTCTT
 20701 CCCCATGGCC CACAACACCG CCTCCACGCT TGAGGCCATG CTTAGAAACG ACACCAACGA
 20761 CCAGTCCTT AACGACTATC TCTCCGCCGC CAACATGCTC TACCCATATAC CCGCCAACGGC
 20821 TACCAACGTG CCCATATCCA TCCCCCTCCCG CAAACTGGGG GCTTCCGCG GCTGGGCTT
 20881 CACGCGCTT AAGACTAAGG AAACCCCATC ACTGGGCTCG GGCTACGACC CTTATTACAC
 20941 CTACTCTGGC TCTATACCT ACCTAGATGG AACCTTTAC CTCAACCCACA CTTTTAAGAA
 21001 GGTGCCATT ACCTTTGACT CTTCTGTCAAG CTGGCCTGGC AATGACCGCC TGCTTACCCC
 21061 CAACGAGTTT GAAATTAAGC GCTCAGTTGA CGGGGAGGGT TACAACGTTG CCCAGTGTAA
 21121 CATGACCAAA GACTGGTTCC TGGTACAAAT GCTAGCTAAC TACAACATTG GCTACCAAGGG
 21181 CTTCTATATC CCAGAGAGCT ACAAGGACCG CAGTACTCC TTCTTTAGAA ACTTCCAGGC
 21241 CATGAGCCGT CAGGTGGTGG ATGATACTAA ATACAAGGAC TACCAACAGG TGGGCATCCT
 21301 ACACCAACAC AAACAACCTG GATTGTTGG CTACCTTGCC CCCACCATGC GCGAAGGACA
 21361 GGCTTACCCCT GCTAACTTCC CCTATCCGCT TATAGGCAAG ACCGCAAGTTG ACAGCATTAC
 21421 CCAGAAAAAG TTCTTTGCG ATCGCACCCCT TTGGCGCATC CCATTCTCCA GTAACTTTAT
 21481 GTCCATGGGC GCACTCACAG ACCTGGGCCA AAACCTTCTC TACGCAACT CCGCCACGC
 21541 GCTAGACATG ACTTTTGAGG TGGATCCCAT GGACGAGCCC ACCCTCTTT ATGTTTTGTT
 21601 TGAAGTCTTT GACGTGGTCC GTGTGCAACCG GCGCACCAGC GGCCTCATCG AAACCGTGTAA
 21661 CCTGCGCACG CCCTTCTCGG CCGGCAACGC CACAACATAA AGAACCAAGC AACATCAACA
 21721 ACAGCTGCCG CCATGGGCTC CAGTAGGCAG GAACTGAAAAG CCATTGTCAA AGATCTTGTT
 21781 TGTGGGCCAT ATTTTTGGG CACCTATGAC AAGCGCTTTC CAGGCTTTGTT TTCTCCACAC
 21841 AAGCTCGCCT CGGCCATAGT CAATAACGCC GGTGCGGAGA CTGGGGCGT ACACTGGATG
 21901 GCCTTGCCT GGAAACCCGCA CTCAAAACAA TGCTACCTCT TTGAGCCCTT TGGCTTTCTT
 21961 GACCAGCGAC TCAAGCAGGT TTACCAAGTT GAGTACGAGT CACTCTGCG CGTAGCGCC
 22021 ATTGCTCTT CCCCCGACCG CTGTATAACG CTGGAAAAGT CCACCCAAAG CGTACAGGGG
 22081 CCCAACTCGG CGGCCCTGG ACTATTCTGC TGCATGTTTC TCCACCCCTT TGCCAACCTGG
 22141 CCCCAAACTC CCATGGATCA CAACCCCAAC ATGAACCTTA TTACCGGGGT ACCCAACTCC
 22201 ATGCTCAACA GTCCCCAGGT ACAGCCCACC CTGCGTCGCA ACCAGGAACA GCTCTACAGC
 22261 TTCCCTGGAGC GCCACTCGCC CTACTTCCGC AGCCACAGTG CGCAGATTAG GAGCGCCACT
 22321 TCTTTTGTC ACTTGAAAAA CATGTAAAAA TAATGTACTA GAGACACTTT CAATAAAGGC
 22381 AAATGCTTTT ATTTGTACAC TCTCGGGTGA TTATTTACCC CCACCCCTTGC CGTCTGCGCC
 22441 GTTTAAAAAT CAAAGGGGTT CTGCCCGCAGA TCGCTATGCG CCACTGGCAG GGACACGTTG
 22501 CGATACTGGT GTTTAGTGT CCACTTAAAC TCAGGCACAA CCATCCGCGG CAGCTGGTG
 22561 AAGTTTCAC TCCACAGGCT GCGCACCATC ACCAACCGCT TTAGCAGGTC GGGCGCCGAT
 22621 ATCTTGAAGT CGCAGTTGGG GCCTCCGCCC TGCGCGCGG AGTTGCGATA CACAGGGTTG
 22681 CAGCACTGGA ACATATCAG CGCCGGGTGG TGCACTGGCTT CCAGCACGCT TTGTCGGAG
 22741 ATCAGATCCG CGTCCAGGCT CTCCCGTTG CTCAGGGCA ACGGAGTCAA CTTTGGTAGC
 22801 TGCTTCCCA AAAAGGGCGC GTGCCAGGC TTGAGTTG ACTCGCACCG TAGTGGCATC
 22861 AAAAGGTGAC CGTGCCCGGT CTGGCGTTA GGATACAGCG CCTGCATAAA AGCCTTGATC
 22921 TGCTAAAAG CCACCTGAGC CTTTGCCT TCAAGAGAAGA ACATGCCGCA AGACTTGCG
 22981 GAAAACGTGAT TGGCCGGACA GGCCCGTGC TGCACGCAGC ACCTTGCGTC GGTGTTGGAG
 23041 ATCTGCACCA CATTGGGCC CCACCGGTTT TTCACGATCT TGGCCCTTGCT AGACTGCTCC

FIG. 8G

64/92

23101 TTCAGCGCGC GCTGCCGTT TTGCTCGTC ACATCCATT CAATCACGTG CTCCCTTATT
 23161 ATCATAATGC TTCCGTGAG ACACCTAACG TCGCCTTCGA TCTCAGCGCA GCGGTGCAGC
 23221 CACAACGCGC AGCCCGTGGG CTCGTGATGC TTGTAGGTCA CCTCTGCAAA CGACTGCAGG
 23281 TACGCCGTCA GGAATCGCCC CATCATCGTC ACAAAAGGTCT TGTGCTGGT GAAGGTCAGC
 23341 TGCAACCCGC GGTGCTCCTC GTTCAGCCAG GTCTTGACATA CGGCCGCCAG AGCTTCCACT
 23401 TGGTCAGGCA GTAGTTTGAA GTTCGCCCTT AGATCGTTAT CCACGTGGTA CTGTCACATC
 23461 AGCGCGCGCG CAGCCTCCAT GCCCTTCTCC CACGCAGACA CGATCGGCAC ACTCAGCGGG
 23521 TTGATCACCAG TAATTTCACT TTCCGCTTCG CTGGGCTCTT CCTCTTCCTC TTGCGTCCGC
 23581 ATACCAACGCG CCACTGGGTC GTCTTCATTC AGCCGCCGA CTGTCGCCCTT ACCTCCCTTG
 23641 CCATGCTTGA TTAGCACCGG TGGGTGCTG AAACCCACCA TTTGTAGCGC CACATCTTCT
 23701 CTTCTTCCTC CGCTGTCCAC GATTACCTCT GGTGATGGCG GGCCTCGGG CTTGGGAGAA
 23761 GGGCGCTCTT TTTCTTCTT GGGCCAAATGCC CCGCCCGAGGT CGATGGCCGC
 23821 GGGCTGGGTG TGCGCGCAC CAGCGCTCT TGTGATGAGT CTTCCCTGTC CTCGGACTCG
 23881 ATACCGCGCC TCATCCGCTT TTTGGGGGC GCCCCGGGAG GCGGCCGCGA CGGGGACGGG
 23941 GACGACACGT CCTCCATGGT TGGGGGACGT CGCGCCGCAC CGCGTCCCGC CTGGGGGGTG
 24001 GTTTCGCGCT GCTCCCTTC CCGACTGGCC ATTTCCCTCT CCTATAGGCA GAAAAGATC
 24061 ATGGAGTCAG TCGAGAAGAA GGACAGCCTA ACCGCCCCCT CTGAGTTGCG CACCACCGCC
 24121 TCCACCGATG CCGCCAACGC GCCTTACCC ACCCTCCGTC AGGCACCCCCC GCTTGAGGAG
 24181 GAGGAAGTGA TTATCGAGCA GGACCCAGGT TTTGTAAGCG AAGACGACGA GGACCGCTCA
 24241 GTACCAACAG AGGATAAAA GCAAGACCGAG GACAACGCGAG AGGCAAAACGA GGAACAAGTC
 24301 GGGCGGGGGG ACGAAAGGCA TGGCGACTAC CTAGATGTGG GAGACGACGT GCTGTTGAAG
 24361 CATCTGCAGC GCCAGTGCAC CATTATCTGC GACCGCTTGC AAGAGCGCAG CGATGTGCC
 24421 CTCGCCATAG CGGATGTCAG CCTTGCTCAC GAACGCCACC TATTCTCACC GCGCGTACCC
 24481 CCCAACCGCC AAAAAAAACGG CACATGCGAG CCCAACCCGC GCCTCAACTT CTACCCCGTA
 24541 TTGCGCTGTC CAGAGGTGCT TGCCACCTAT CACATCTTT TCCAAAATCTG CAAGATAACCC
 24601 CTATCCGTGC GTGCCAACCG CAGCCGAGCG GACAAGCAGC TGGCTTGC GCAAGGGCGCT
 24661 GTCATACCTG ATATCGCTC GCTCAACGAA GTGCCAAAAA TCTTGTAGGG TCTTGGACGC
 24721 GACGAGAAGC GCGCGGCAAA CGCTCTGCAA CAGGAAAACA GCGAAAATGA AAGTCACTCT
 24781 GGAGTGTGTTGG TGGAACTCGA GGGTACAAC CGCGCGCTAG CGTACTAAA AGCAGCATC
 24841 GAGGTCACCC ACTTTGCCA CCCGGCACTT AACCTACCCC CCAAGTCAT GAGCACAGTC
 24901 ATGAGTGAGC TGATCGTGC CCGTGCAG CCCCCTGGAGA GGGATGCAAA TTGCAAGAA
 24961 CAAACAGAGG AGGGCCTACC CGCAGTTGGC GACGAGCAGC TAGCGCGCTG GCTTCAAC
 25021 CGCGAGCCTG CCGACTTGGA GGAGCGACGC AAACATATGA TGCCCGAGT GCTCGTTACC
 25081 GTGGAGCTTG AGTGCATGCA CGGGTCTTT GCTGACCCCG AGATGCAGCG CAAGCTAGAG
 25141 GAAACATTCG ACTACACCTT TCGACAGGGC TACGTACGCC AGGCTGCAA GATCTCCAAC
 25201 GTGGAGCTCT GCAACCTGGT CTCCTACCTT GGAATTGGT ACGAAAACCG CCTTGGC
 25261 AACGTGCTTC ATTCCACGCT CAAGGGCGAG CGCGCGCCGG ACTACGTCCG CGACTGCGTT
 25321 TACTTATTT TATGTCACAC CTGGCAGACG GCCATGGGGC TTGGCAGCA GTGCTTGGAG
 25381 GAGTGCACC TCAAGGAGCT GCAGAAAATG CTAAGGCAAA ACTTGAGGA CCTATGGAC
 25441 GCCTTCACG AGCGCTCCGT GGCGCGCAC CTGGCGGACA TCATTTCCC CGAACGCC
 25501 CTTAAAACCC TGCAACAGGG TCTGCCAGAC TTCACCAAGTC AAAGCATGTT GCAGAACTTT
 25561 AGGAACCTTA TCCTAGAGCG CTCAGGAATC TTGCCCCCA CCTGCTGTGC ACTTCCTAGC
 25621 GACTTTGTGC CCATTAAGTA CCGCGAATGC CCTCCGCCG TTTGGGCCA CTGCTACCTT
 25681 CTGCAAGCTAG CCAACTACCT TGCCCTACAC TCTGACATAA TGGAAAGACGT GAGCGGTGAC
 25741 GGTCTACTGG AGTGTCACTG TCGCTGCAAC CTATGCACCC CGCACCGCTC CCTGGTTTG
 25801 AATTGCGAGC TGCTTAACGA AAGTCAAATT ATCGGTACCT TTGAGCTGCA GGGTCCCTCG
 25861 CCTGACGAAA AGTCCGGCGC TCCGGGGTTG AAACTCACCT CGGGGCTGTG GACGTGCGCT
 25921 TACCTTCGCA AATTTGTACC TGAGGACTAC CACGCCACAG AGATTAGGTT CTACGAAGAC
 25981 CAATCCCGCC CGCCAAATGC GGAGCTTACG GCCTGCGTCA TTACCCAGGG CCACATTCTT
 26041 GGCCAATTGC AAGCCATCAA CAAAGCCCGC CAAGAGTTTC TGCTACGAAA GGGACGGGG
 26101 GTTACTTGG ACCCCCCAGTC CGGGCAGGAG CTCAACCCAA TCCCCCGCC GCCGCAGCCC
 26161 TATCAGCAGC AGCCGCGGGC CCTTGCTTCC CAGGATGGCA CCCAAAAAGA AGCTGCAGCT
 26221 GCGGCCGCCA CCCACGGACG AGGAGGAATA CTGGGACAGT CAGGAGGAGG AGGTTTTGG
 26281 CGAGGAGGAG GAGGACATGA TGGAAAGACTG GGAGAGCCTA GACGAGGAAG CTTCCGAGGT
 26341 CGAAAGAGGTG TCAGACGAAA CACCGTCACC CTGGCTGCA TTCCCCCTCGC CGGC
 GCGCCCCCA

FIG. 8H

65/92

26401 GAAATCGGCA ACCGGTTCCA GCATGGCTAC AACCTCCGCT CCTCAGGGCGC CGCCGGCACT
 26461 GCCCCTTCGC CGACCCAACC GTAGATGGGA CACCACTGGA ACCAGGGCCG GTAAGTCAA
 26521 GCAGCCGCCG CCGTTAGCCC AAGAGCAACA ACAGCGCAA GGCTACCGCT CATGGCGCGG
 26581 GCACAAGAAC GCCATAGTTG CTTGTTGCA AGACTGTGGG GGCAACATCT CCTTCGCCCG
 26641 CCGCTTCTT CTCTACCATC ACGGCGTGGC CTTCCCCCGT AACATCCTGC ATTACTACCG
 26701 TCATCTCTAC AGCCCCACT GCACCCGGG CAGCGGAGC GGCAGCAACA GCAGCGGCA
 26761 CACAGAAGCA AAGGCAGCCG GATAGCAAGA CTCTGACAAA GCCAAGAAA TCCACAGCGG
 26821 CGGCAGCAGC AGGAGGAGGA GCGCTCGTC TGGCGCCCAA CGAACCCGTA TCGACCCGCG
 26881 AGCTTAGAAA CAGGATTTT CCCACTCTGT ATGCTATATT TCAACAGAGC AGGGGCAAG
 26941 AACAAAGAGCT GAAAATAAAA AACAGGTCTC TGCATGCCCT CACCCGCAGC TGCCCTGTATC
 27001 ACAAAAGCGA AGATCAGCTT CGGCCACGC TGGAAAGACGC GGAGGCTCTC TTCAGTAAT
 27061 ACTCGCGCCT GACTCTTAAG GACTAGTTTGC CGCCCTTTC TCAAATTAA CGCGGAAAAC
 27121 TACGTCATCT CCAGCGGCCA CACCCGGCGC CAGCACCTGT CGTCAGCGCC ATTATGAGCA
 27181 AGGAAATTCC CACGCCCTAC ATGTTGAGTT ACCAGCCACA AATGGGACTT GCGGCTGGAG
 27241 CTGCCCCAAGA CTACTCAACC CGAATAAAACT CATGAGCGC GGGACCCAC ATGATATCCC
 27301 GGGTCAACGG AATCCGGCC CACCGAAACC GAATTCTCTT GGAACAGGCG GCTATTACCA
 27361 CCACACCTCG TAATAACCTT AATCCCCGTA GTTGGCCCG TGCCCTGGTG TACCAAGAAA
 27421 GTCCCGCTCC CACCACTGTG GTACTTCCCA GAGACGCCA GGGCGAAGTT CAGATGACTA
 27481 ACTCAGGGGC GCAGCTTGC GGCAGCTTTC GTCACAGGGT GCGGTGCGCC GGGCAGGGTA
 27541 TAACTCACCT GACAATCAGA GGGCGAGGTA TTCAGCTCAA CGACGAGTCG GTGAGCTCCT
 27601 CGCTTGGTCT CGTCCGGAC GGGACATTTC AGATCGGCGG CGCCGGCGT CCTTCATTCA
 27661 CGCCTCGTCA GGCAATCTTA ACTCTGCAGA CCTCGTCCTC TGAGCCGCGC TCTGGAGGCA
 27721 TTGGAACTCT GCAATTATTGAGTT GAGGAGTTTG TGCCATCGGT CTACTTTAAC CCCTTCTCGG
 27781 GACCTCCCGG CCACTATCCG GATCAATTAA TTCCCTAACT TGACCGGTA AAGGACTCGG
 27841 CGGACGGCTA CGACTGAATG TTAAGTGGAG AGGCAGAGCA ACTGCCCTG AACACCTGG
 27901 TCCACTGTG CGGCCACAAG TGCTTGCCTC GCGACTCCCG TGAGTTTG TGACTTTGAAT
 27961 TGCCCGAGGA TCATATCGAG GGCCCGCGC ACGGCGTCCG GCTTACCGCC CAGGGAGAGC
 28021 TTGCCCCGTAG CCTGATTCCG GAGTTTACCC AGCGCCCCCT GCTAGTTGAG CGGGACAGGG
 28081 GACCCGTGT TCTCACTGTG ATTTGCAACT GTCCTAACCT TGGATTACAT CAAGATCTT
 28141 GTTGCCATCT CTGTGCTGAG TATAATAAAAT ACAGAAATTAA ATATATAACTG GGGCTCTAT
 28201 CGCCATCTCG TAAACGCCAC CGTCTTCACC CGCCCAAGCA AACCAAGGCG AACCTTACCT
 28261 GGTACTTTA ACATCTCTCC CTCTGTGATT TACAACAGTT TCAACCCAGA CGGAGTGAGT
 28321 CTACGAGAGA ACCTCTCCGA GCTCAGCTAC TCCATCAGAA AAAACACCAC CCTCCTTACC
 28381 TGCCGGAAAC GTACGAGTGC GTCACCGGCC GCTGCACCCAC ACCTACCGCC TGACCGTAAA
 28441 CCAGACTTTT TCCGGACAGA CCTCAATAAC TCTGTTTACCG AGAACAGGAG GTGAGCTTAG
 28501 AAAACCCCTTA GGGTATTAGG CCAAAGGCGC AGCTACTGTG GGGTTATG ACAAATTCAAG
 28561 CAACTCTACG GGCTATTCTA ATTCAAGGTTT CTCTAGAAC GGGGTGGGG TTATTCTCTG
 28621 TCTTGTGATT CTCTTTATTC TTATCAACTAC GCTTCTCTGC CTAAGGCTCG CGCCCTGCTG
 28681 TGTGACATT TGCACTTATT GTCAGCTTT TAAACGCTGG GGTGCCACC CAAGATGATT
 28741 AGGTACATAA TCCTAGGTTT ACTCACCCCTT GCGTCAGCCC ACGGTACCCAC CAAAAGGTG
 28801 GATTAAAG AGCCAGCTG TAATGTTACA TTCGAGCTG AAGCTAATGA GTGCACCACT
 28861 CTTATAAAAT GCACCAACAGA ACATGAAAAG CTGCTTATTC GCCACAAAAA CAAAATTGGC
 28921 AAGTATGCTG TTTATGCTAT TTGGCAGCCA GGTGACACTA CAGAGTATAA TGTTACAGTT
 28981 TTCCAGGGTA AAAGTCATAA AACTTTATG TATACTTTTC CATTATGAA AATGTGCGAC
 29041 ATTACCATGT ACATGAGCAA ACAGTATAAG TTGTGGCCCA CACAAAATTG TGTGGAAAAC
 29101 ACTGGCACTT TCTGCTGCAC TGCTATGCTA ATTACAGTGC TCGCTTGGT CTGTACCCCTA
 29161 CTCTATATTA AATACAAAAG CAGACGCAGC TTTATTGAGG AAAAGAAAAT GCCTTAATT
 29221 ACTAAGTTAC AAAGCTAATG TCACCACTAA CTGCTTTACT CGCTGCTTGC AAAACAAAATT
 29281 CAAAAAGTTA GCATTATAAT TAGAATAGGA TTAAACCCC CGGTCACTTT CCTGCTCAAT
 29341 ACCATTCCCC TGAACAAATTG ACTCTATGTG GGATATGCTC CAGCGCTACA ACCTTGAAGT
 29401 CAGGCTTCCT GGATGTCAGC ATCTGACTTT GGGCAGCACC TGICCCGGG ATTGTTCTCA
 29461 GTCCAACTAC AGCGACCCAC CCTAACAGAG ATGACCAACA CAACCAACGC GGCCGCCGCT
 29521 ACCGGACTTA CATCTACCC AAATACACCC CAAGTTCTG CCTTGTCAA TAATGGGAT
 29581 AACTTGGGCA TGTGGTGGTT CTCCATAGCG CTTATGTTG TATGCCCTTAT TATTATGTGG
 29641 CTCATCTGCT GCCTAAAGCG CAAACGGCC CGACCAACCA TCTATAGTCC CATCATTGTG

FIG. 81

29701 CTACACCCAA ACAATGATGG AATCCATAGA TTGGACGGAC TGAAACACAT GTTCTTTCT
 29761 CTTACAGTAT GATTAATGA GACATGATT CTCGAGTTT TATATTACTG ACCCTTGTG
 29821 CGCTTTTGTG TCGGTGCTCC ACATGGCTG CGGTTTCTCA CATCGAAGTA GACTGCATT
 29881 CAGCCTTCAC AGTCTATTG CTTTACGGAT TTGTCACCC CACGCTCATC TGCAGCCTCA
 29941 TCACTGTGGT CATGCCCTT ATCCAGTGCA TTGACTGGGT CTGTGTGCGC TTTGCATATC
 30001 TCAGACACCA TCCCCAGTAC AGGGACAGGA CTATAGCTGA GCTTCTTAGA ATTCTTTAAT
 30061 TATGAAATT ACTGTGACTT TTCTGCTGAT TATTTGCACC CTATCTGCGT TTTGTTCCCC
 30121 GACCTCCAAG CCTCAAAGAC ATATATCATG CAGATTCACT CGTATATGGA ATATTCAG
 30181 TTGCTACAAT GAAAAAAGCG ATCTTCCGA AGCCTGGTTA TATGCAATCA TCTCTGTTAT
 30241 GGTGTTCTGC AGTACCATCT TAGCCCTAGC TATATATCCC TACCTTGACA TTGGCTGGAA
 30301 ACGAATAGAT GCCATGAACC ACCCAACTTT CCCCGCGCCC GCTATGCTTC CACTGCAACA
 30361 AGTTGTTGCC GGCGGCTTGT TCCCAGCCAA TCAGCCTCGC CCCACTCTC CCACCCCCAC
 30421 TGAATCAGC TACTTTAAC TAAACAGGAGG AGATGACTGA CACCCTAGAT CTAGAAATGG
 30481 ACGGAATTAT TACAGAGCAG CGCCTGCTAG AAGAGCAGC GGCAGCGGC GAGCAACAGC
 30541 GCATGAATCA AGAGCTCCAA GACATGGTTA ACTTGCACCA GTGCAAAGG GGTATCTTT
 30601 GTCTGGTAA GCAGGGCCAA GTCACCTACG ACAGTAATAC CACCGGACAC CGCCTTAGCT
 30661 ACAAGTTGCC AACCAAGCGT CAGAAATTGG TGGTCATGGT GGGAGAAAAG CCCATTACCA
 30721 TAACTCAGCA CTCGGTAGAA ACCGAAGGCT GCATTCACTC ACCTTGTCAA GGACCTGAGG
 30781 ATCTCTGCAC CCTTATTAG ACCCTGTGCG GTCTCAAAGA TCTTATTCCC TTTAACTAAT
 30841 AAAAAAAAT AATAAAGCAT CACTTACTTA AAATCAGTTA GCAAATTCT GTCCAGTTA
 30901 TTCAGCAGCA CCTCCTGCC CTCCTCCCAG CTCTGGTATT GCAGCTTCCT CCTGGCTGCA
 30961 AACTTCTCC ACAATCTAAA TGGAATGTCA GTTCCCTCCCT GTTCCCTGTC ATCCGCACCC
 31021 ACTATCTCA TGGTGTGCC GATGAAGCGC GCAAGACCGT CTGAAGATAC CTTCAACCCC
 31081 GTGTATCCAT ATGACACCGA AACCGGTCTC CCAAATGTGC CTTTCTTAC TCCCTCCCTT
 31141 GTATCCCCA ATGGGTTTCA AGAGACTCCC CTCGGGGTAC TCTCTTGCCT CCTATCCGA
 31201 CCTCTAGTT CCTCCAATGG CATGGTGC CTCAAAATGG GCAACGGCCT CTCTCTGGAC
 31261 GAGGCCGGCA ACCTTACCTC CCAAATGTA ACCACTGTGA GCCCACCTCT CAAAAAAACC
 31321 AAGTAAACAA TAAACCTGGA AATATCTGCA CCCCTCACAG TTACCTCAGA AGCCCTAACT
 31381 GTGGCTGCCG CGCACCTC AATGGTCGCG GGCAACACAC TCACCATGCA ATCACAGGCC
 31441 CGCTAACCG TGACGACTC CAAACTTAGC ATTGCCACCC AAGGACCCCT CACAGTGTCA
 31501 GAAGGAAAGC TAGCCCTGCA AACATCAGGC CCCCTCACCA CCACCGATAG CAGTACCCCT
 31561 ACTATCACTG CCTCACCCCC TCTAACTACT GCCACTGGTA GCTTGGGCAT TGACTTGAAA
 31621 GAGCCCATTT ATACACAAAA TGAAAACAT GGACTAAAGT ACGGGGCTCC TTGCAATGTA
 31681 ACAGACGACC TAAACACTT GACCGTAGCA ACTGGTCCAG GTGTGACTAT TAATAATACT
 31741 TCCTTGCAAA CTAAAGTTAC TGGAGCCTTG GTTTTGATT CACAAGGCAA TATGCAACTT
 31801 AATGTAGCAG GAGGACTAAG GATTGATTCT CAAAACAGAC GCCTTATACT TGATGTTAGT
 31861 TATCCGTTTG ATGCTAAAA CCAACTAAAT CTAAGACTG GACAGGGCCC TCTTTTTATA
 31921 AACTCAGCCC ACAACTTGGA TATTAACATC ACAAAAGGCC TTTACTTGT TACAGCTTC
 31981 AACAAATTCCA AAAAGCTGA GGTAAACCTA AGCACTGCA AGGGGGTGTAT GTTGTACGCT
 32041 ACAGCCATAG CCATTAATGC AGGAGATGGG CTTGAATTG GTTCACCTAA TGCACCAAAAC
 32101 ACAATCCCC TCAAAACAAA AATTGGCCAT GGCTAGAAT TTGATTCAA CAAGGCTATG
 32161 GTTCTAACAC TAGGAACCTG CCTTAGTTT GACAGCACAG GTGCCATTAC AGTAGGAAAC
 32221 AAAATAATG ATAAGCTAAC TTTGTGGACC ACACCAGCTC CATCTCTAA CTGTAGACTA
 32281 AATGCAGAGA AAGATGCTAA ACTCACTTTG GTCTTAACAA AATGTGGCAG TCAAATACTT
 32341 GCTACAGTTT CAGTTTGGC TGTTAAAGGC AGTTTGGCTC CAATATCTGG AACAGTTCAA
 32401 AGTGTCTATC TTATTATAAG ATTTGACGAA AATGGAGTGC TACTAAACAA TTCCCTCC
 32461 GACCCAGAAAT ATTGGAAACCTT TAGAAATGGA GATCTTACTG AAGGCACAGC CTATACAAAC
 32521 GCTGTTGGAT TTATGCCAA CCTATCAGCT TATCCAAAAT CTCACCGTAA AACTGCCAAA
 32581 AGTAACATTG TCAGTCAAGT TTACTTAAAC GGAGACAAAA CTAACACTGT AACACTAAC
 32641 ATTACACTAA ACGGTACACA GGAAACAGGA GACACAACTC CAAGTGCATA CTCTATGTCA
 32701 TTTCATGGG ACTGGTCTGG CCACAACTAC ATTAATGAAA TATTGCCAC ATCCCTTAC
 32761 ACTTTTCTATC ACATTGCCA AGAATAAAGA ATCGTTTGTG TTATGTTCA ACGTGTTTAT
 32821 TTTCAATTG CAGAAAATTG CAAAGTCAATT TTCATTCACT AGTATAGCCC CACCACCA
 32881 TAGCTTATAC AGATCACCGT ACCTTAATCA AACTCACAGA ACCCTAGTAT TCAACCTGCC
 32941 ACCTCCCTCC CAACACACAG AGTACACAGT CCTTCTCCC CGGCTGGCCT TAAAAGCAT

67/92

33001 CATATCATGG GTAACAGACA TATTCTTAGG TGTTATATTG CACACGGTTT CCTGTCGAGC
 33061 CAAACGCTCA TCAGTGATAT TAATAAACTC CCCGGGCAGC TCACTTAAGT TCATGTCGCT
 33121 GTCCAGCTGC TGAGCCACAG GCTGCTGTCC AACTTGCAGG TGCTTAACGG CGGGCGAAGG
 33181 AGAAGTCCAC GCCTACATGG GGGTAGAGTC ATAATCGTC ATCAGGATAG GGCGGTGGTG
 33241 CTGCAGCAGC GCGCGAATAA ACTGCTGCCG CCGCCGCTCC GTCCTGCAGG AATACAACAT
 33301 GGCAGTGGTC TCCTCAGCGA TGATTGCGAC CGCCCGCAGC ATAAGGCGCC TTGTCCCTCG
 33361 GGCACAGCAG CGCACCCCTGA TCTCACTTAA ATCAGCACAG TAACTGCAGC ACAGCACCA
 33421 AATATTGTTA AAAATCCCAC AGTGCAAGGC GCTGTATCCA AAGCTCATGG CGGGGACCA
 33481 AGAACCCACG TGGCCATCAT ACCACAAGCG CAGGTAGATT AAGTGGCGAC CCCTCATAAA
 33541 CACGCTGGAC ATAAACATTA CCTCTTTTGG CATGTTGTAA TTCACACCTC CCCGGTACCA
 33601 TATAAACCTC TGATTAACAA TGGGCCCATC CACCAACATC CTAAACCAGC TGGCCAAA
 33661 CTGCCGCCG GCTATACACT GCAGGGAAACC GGGACTGAA CAATGACAGT GGAGAGCCCA
 33721 GGACTCGTAA CCATGGATCA TCATGCTCGT CATGATATCA ATGTTGGCAC AACACAGGC
 33781 CACGTGCATA CACTTCCTCA GGATTACAAG CTCCCTCCCG GTTAAACCA TATCCCAGGG
 33841 AACAAACCCAT TCCGTAAATCA GCGTAAATCC CACACTGCAG GGAAGACCTC GCACGTAAC
 33901 CACGTTGTGC ATTGTCAAAG TGTTACATTG GGGCAGCAGC GGATGATCCT CCAGTATGGT
 33961 AGCGGGGTT TCTGTCTCAA AAGGAGGTAG ACGATCCCTA CTGTACGGAG TGCGCCGAGA
 34021 CAACCGAGAT CGTGTGGTC GTAGTGTCA GCAAATGGA ACGCCGGACG TAGTCATATT
 34081 TCCTGAAGCA AAACCAGGTG CGGGCGTGAC AAACAGATCT GCGTCTCCGG TCTCGCCGCT
 34141 TAGATCGCTC TGTGTAGTAG TTGTAGTATA TCCACTCTCT CAAAGCATCC AGGCGCCCC
 34201 TGGCTTCGGG TTCTATGTA ACTCCCTCAT GCGCCGCTGC CCTGATAACA TCCACCACCG
 34261 CAGAATAAGC CACACCCAGC CAACCTACAC ATTGCTTCG CGAGTCACAC ACGGGAGGGAG
 34321 CGGGAAAGAGC TGGAAGAACC ATGTTTTTTT TTTTATTCCA AAAGATTATC CAAAACCTCA
 34381 AAATGAAGAT CTATTAAGTG AACCGCGCTCC CTCGGTGG CGTGGTCAAA CTCTACAGCC
 34441 AAAGAACAGA TAATGGCATT TGTAAGATGT TGCACAAATGG CTTCCAAAAG CAAACAGGCC
 34501 CTCACGTCCA AGTGGACGTA AAGGCTAAAC CTTTCAGGGT GAATCTCCTC TATAAACATT
 34561 CCAGCACCTT CAACCATGCG CAAATAATTG TCATCTCGG ACCTTCTCAA TATATCTCTA
 34621 AGCAAAATCC GAATATTAAAG TCCGGCCATT GTAAAATCT GCTCCAGAGC GCCCTCCACC
 34681 TTACGCTCA AGCAGCGAAT CATGATTGCA AAAATTCAAGG TTCCCTCACAG ACCTGTATAA
 34741 GATTCAAAAG CGGAACATTA ACAAAAATAC CGCGATCCCG TAGGTCCTT CGCAGGGCCA
 34801 GCTGAACATA ATCGTGCAGG TCTGCACGGA CCAGCGCGGC CACTTCCCCG CCAGGAACCT
 34861 TGACAAAAGA ACCCACACTG ATTATGACAC GCATACTCGG AGCTATGCTA ACCAGCGTAG
 34921 CCCCGATGTA AGCTTTGTG CATGGCGGC GATATAAAAT GCAAGGTGCT GCTCAAAAAA
 34981 TCAGGCAAAG CCTCGCGCAA AAAAGAAAGC ACATCGTAGT CATGCTCATG CAGATAAAAGG
 35041 CAGGTAAGCT CGGAAACCAC CACAGAAAAA GACACCATTT TTCTCTCAA CATGTCGCG
 35101 GGTTCCTGCA TAAACACAAA ATAAAATAAC AAAAAGACAT TAAACACATTA GAAGCCTGTC
 35161 TTACAAACAGG AAAAACAAAC CTTATAAGCA TAAGACGGAC TACGCCATG CGGGCGTGAC
 35221 CGTAAAAAAA CTGGTCACCG TGATTAAGG GCACCCACCGA CAGCTCCTCG GTCATGTCCG
 35281 GAGTCATAAT GTAAGACTCG GTAAACACAT CAGGTTGATT CATCGGTAG TGCTAAAAG
 35341 CGACCGAAAT AGCCCGGGGG AATACATACC CGCAGGGCTA GAGACAAACAT TACAGCCCC
 35401 ATAGGAGGT AAACAAAATT AATAGGAGAG AAAAACACAT AAACACCTGA AAAACCTCC
 35461 TGCCTAGGC AAATAGCACC CTCCCGCTCC AGAACAAACAT ACAGCGCTTC ACAGCGGCAG
 35521 CCTAACAGTC AGCCTTACCA GTAAAAAGA AAACCTATTA AAAAACACC ACTCGACACG
 35581 GCACCAAGCTC AATCAGTCAC AGTGTAAAAA AGGGCCAAGT GCAGAGCGAG TATATATAGG
 35641 ACTAAAAAAAT GACGTAACGG TTAAAGTCCA CAAAAAACAC CCAGAAAACC GCACGCGAAC
 35701 CTACGCCCAG AAACGAAAGC CAAAAAACCC ACAACTTCCT CAAATCGTCA CTTCCGTTT
 35761 CCCACGTTAC GTAACCTCCC ATTTTAAGAA AACTACAATT CCCAACACAT ACAAGTTACT
 35821 CGGCCCTAAA ACCTACGTCA CCCGCCCGT TCCCACGCC CGCGCCACGT CACAAACTCC
 35881 ACCCCCTCAT TATCATATTG GCTTCATACCA AAAATAAGT ATATTATTGA TGATG

FIG. 8K

68/92

Structure of the Ad6 Genome

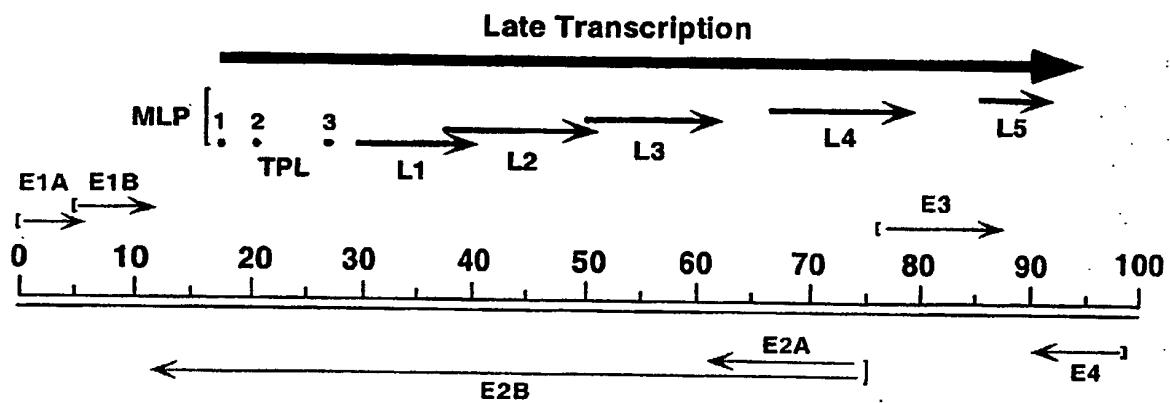


FIG. 9

69/92

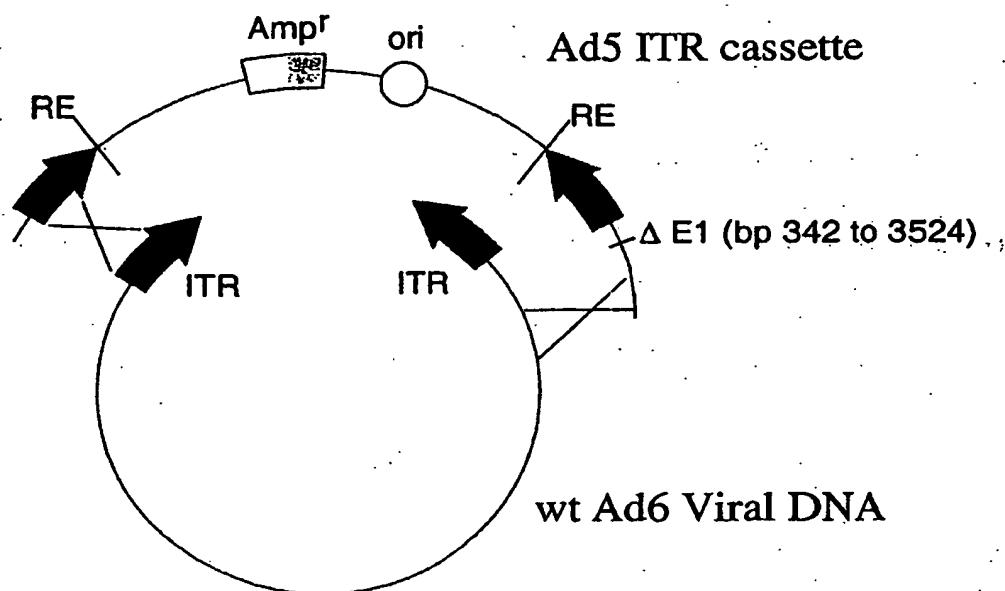


FIG. 10

70/92

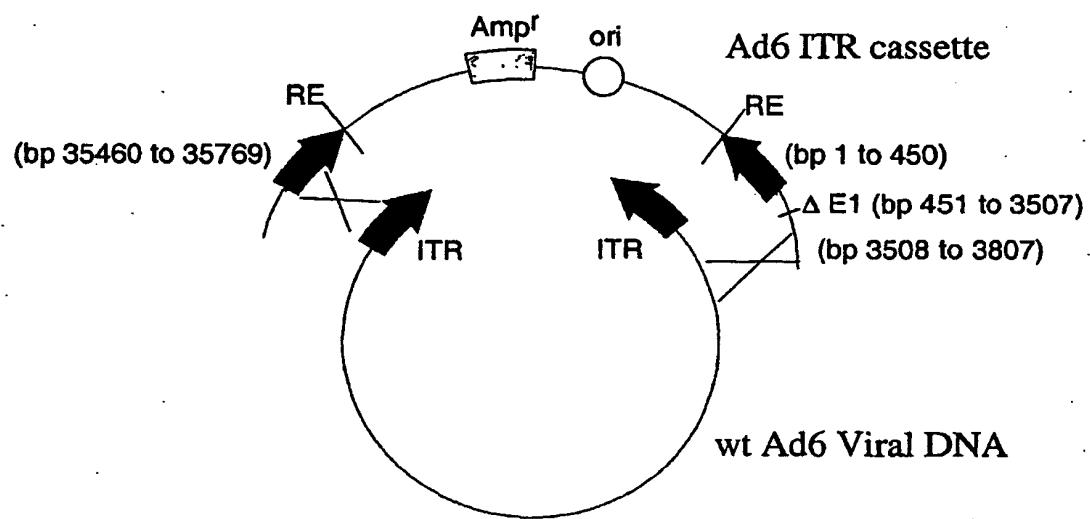
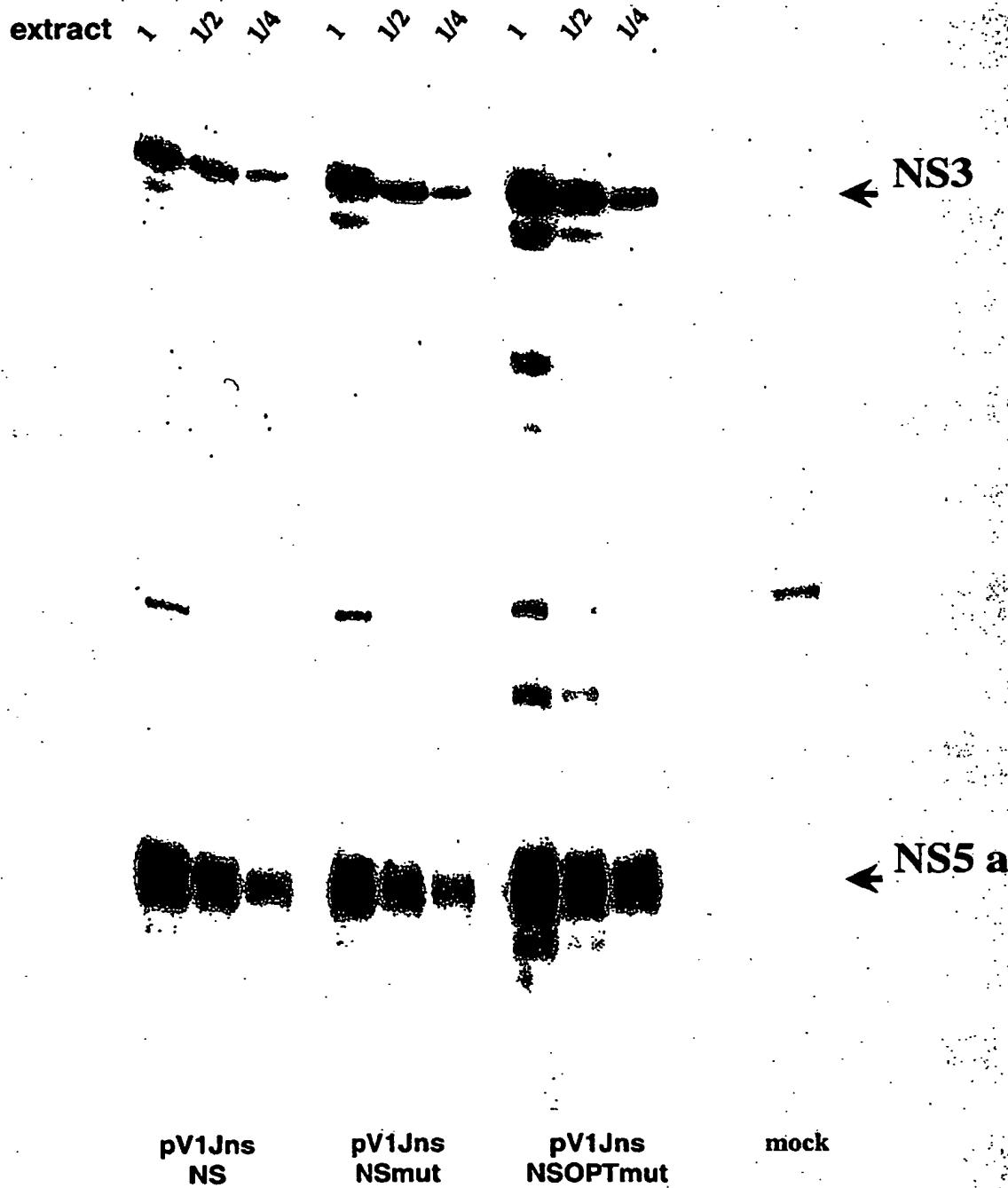


FIG. 11

71/92



Western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing the different HCV NS cassettes. Mature NS3 and NSSA products were detected with specific antibodies.

FIG. 12

72/92

mouse	Pep pool								
	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSO	
pV1jns-NS	#31	41	135	19	44	25	17	137	8
	#32	121	783	77	144	13	22	604	4
	#33	8	32	3	11	6	6	43	3
	#34	16	139	13	47	31	25	151	2
	#35	21	101	40	32	21	20	75	1
	#36	18	26	24	25	5	7	29	6
	#37	19	73	15	39	8	20	49	2
	#38	133	575	74	345	75	63	515	5
	#39	40	183	10	85	14	9	148	2
	#40	66	465	29	111	15	16	189	0
Geomean		33	146	21	57	15	16	123	na

mouse	Pep pool								
	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSO	
pV1jns-NSmut	#41	39	293	58	187	5	4	248	1
	#42	21	220	46	107	26	10	189	4
	#43	76	134	12	78	8	6	144	2
	#44	30	45	20	52	4	8	40	4
	#45	36	100	17	56	4	6	116	3
	#46	67	172	16	138	8	9	145	3
	#47	34	131	28	38	9	5	118	1
	#48	55	316	43	107	9	7	277	5
	#49	6	131	5	25	4	1	91	0
	#50	13	93	11	11	5	1	76	1
Geomean		30	142	20	61	7	5	126	na

mouse	Pep pool								
	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSO	
V1jns-NSOPTmut	#51	53	409	34	84	11	25	271	4
	#52	140	660	65	276	23	36	377	2
	#53	58	553	48	105	23	18	564	1
	#54	50	105	35	134	10	16	80	2
	#55	14	80	11	35	4	7	91	6
	#56	14	342	30	101	23	14	207	1
	#57	63	325	66	239	17	24	123	1
	#58	75	542	66	168	127	93	191	0
	#59	65	468	40	124	18	23	344	4
	#60	27	142	48	16	7	8	77	0
Geomean		45	295	40	99	16	20	188	na

IFN γ ELISPOT on splenocytes from C57black6 mice immunized with two injections of 25 μ g DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10 6 PBMC.

FIG. 13A

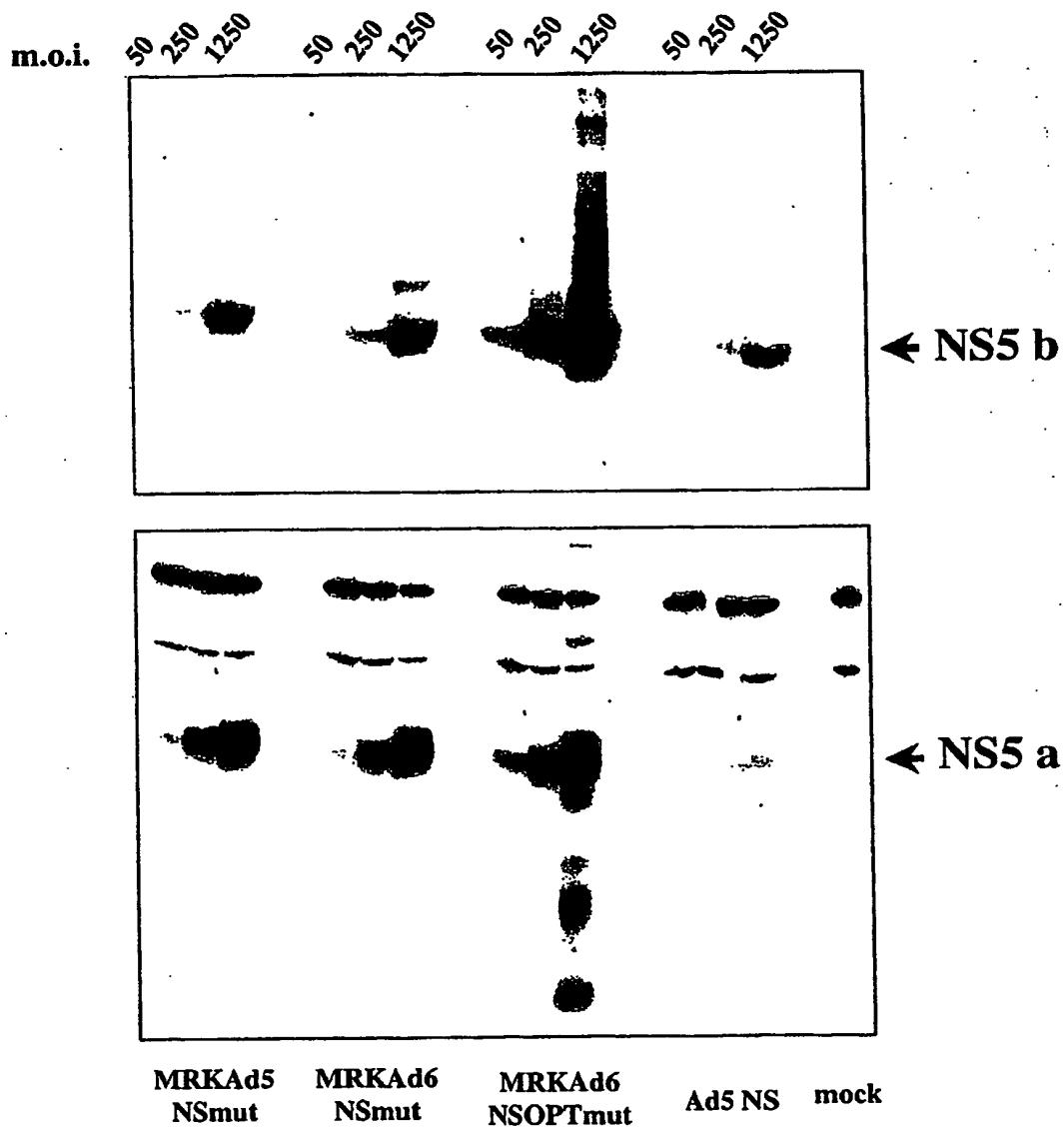
73/92

Pep pool							
mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
pV1jns-NS	#51	219	699	634	486	487	264
	#52	67	302	347	167	111	87
	#53	59	460	400	246	244	136
	#54	139	817	685	236	547	223
	#55	96	904	542	277	256	337
	#56	225	603	686	156	350	240
	#57	44	288	211	148	100	141
	#58	37	262	221	53	58	62
	#59	131	975	928	159	305	284
	#60	93	475	464	77	206	113
geo mean		111	579	512	201	266	189
							20
pV1jns-NSmut	mouse	Pep pool					
	#61	72	840	515	219	278	249
	#62	294	1881	1266	365	434	411
	#63	73	415	422	103	141	99
	#64	66	824	486	175	162	144
	#66	24	313	168	53	47	42
	#67	15	230	253	94	25	39
	#68	53	354	252	89	101	86
	#69	271	895	909	518	322	285
	#70	417	1303	1186	468	557	267
geo mean		143	784	606	232	230	180
							30
V1jns-NSOPTmut	mouse	Pep pool					
	#71	206	944	890	342	207	397
	#72	393	1655	1151	575	626	401
	#73	123	522	515	319	223	198
	#74	500	1414	1419	878	1035	1122
	#75	286	812	873	382	543	267
	#76	224	1143	942	218	420	281
	#77	95	643	630	169	385	218
	#78	401	1302	1068	538	608	623
	#79	108	1190	914	199	265	215
geo mean		209	941	854	331	406	329
							24

IFNy ELispot on splenocytes from BalbC mice immunized with two injections of 50 μ g DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 13B

74/92



Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (m.o.i.; indicated at the top) with Adenovectors expressing the different HCV NS cassettes. Mature NS5B and NS5A products were detected with specific antibodies.

FIG. 14

75/92

mouse	Pep pool							
	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)	DMSO	
Ad5-NS	#1	14	492	9	27	10	554	7
	#2	8	440	2	26	5	438	0
	#3	12	92	5	12	7	73	4
	#4	16	388	6	40	6	228	2
	#6	8	210	4	31	3	238	3
	#7	7	133	13	16	0	128	9
	#8	11	342	25	55	22	267	12
	#9	5	345	0	45	5	285	3
	#10	22	888	3	65	25	799	1
	Geomean	10	305	na	31	na	269	na
MRKAd5-NSmut	Pep pool							
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)	DMSO
	#11	14	1009	13	75	7	751	6
	#12	15	695	3	39	9	552	1
	#13	12	389	4	20	7	352	3
	#14	7	459	6	50	1	274	1
	#15	5	549	3	22	6	485	0
	#16	10	631	1	6	4	600	3
	#17	5	257	3	9	1	245	3
	#18	13	659	6	43	7	555	1
	#19	12	758	1	37	5	669	0
	#20	22	1380	5	163	8	1003	4
	Geomean	10	615	3	31	4	504	na
MRKAd6-NSmut	Pep pool							
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)	DMSO
	#21	6	584	5	27	4	491	2
	#22	6	231	3	12	3	235	0
	#23	8	482	1	18	1	511	0
	#24	14	1120	6	38	10	1004	5
	#25	1	311	3	9	0	382	1
	#26	29	903	3	60	5	751	5
	#27	35	1573	4	40	4	1277	4
	#28	7	406	5	15	1	443	3
	#29	4	461	3	12	3	515	3
	Geomean	8	567	3	21	na	554	na

IFN γ ELISPOT on splenocytes from C57black6 mice immunized with two injections of 10⁹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 15

76/92

Pep pools	Ad5-NS 10^{10} vp/dose		
	96074	134T	063Q
<i>F (NS3p)</i>	374	11	74
<i>G (NS3h)</i>	359	1070	1455
<i>H (NS4)</i>	376	30	64
<i>I (NS5a)</i>	240	40	63
<i>L (NS5b)</i>	226	29	121
<i>M (NS5b)</i>	511	23	35
<i>DMSO</i>	128	3	31

Pep pools	MRK Ad6-NSmut 10^{10} vp/dose		
	S207	035Q	057Q
<i>F (NS3p)</i>	363	382	150
<i>G (NS3h)</i>	180	316	119
<i>H (NS4)</i>	126	113	62
<i>I (NS5a)</i>	1780	688	114
<i>L (NS5b)</i>	447	111	81
<i>M (NS5b)</i>	153	38	16
<i>DMSO</i>	9	6	9

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16A

77/92

Pep pools	MRK Ad5-NSmut 10^{10} vp/dose		
	S201	075Q	137Q
<i>F</i> (<i>NS3p</i>)	928	69	254
<i>G</i> (<i>NS3h</i>)	317	436	98
<i>H</i> (<i>NS4</i>)	56	101	45
<i>I</i> (<i>NS5a</i>)	1530	1100	413
<i>L</i> (<i>NS5b</i>)	149	23	92
<i>M</i> (<i>NS5b</i>)	398	32	80
<i>DMSO</i>	29	6	29

Pep pools	MRK Ad6-NSOPTmut 10^{10} vp/dose		
	98D209	106Q	113Q
<i>F</i> (<i>NS3p</i>)	3110	263	404
<i>G</i> (<i>NS3h</i>)	2115	642	1008
<i>H</i> (<i>NS4</i>)	373	72	19
<i>I</i> (<i>NS5a</i>)	103	37	347
<i>L</i> (<i>NS5b</i>)	149	22	10
<i>M</i> (<i>NS5b</i>)	314	428	19
<i>DMSO</i>	0	1	3

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16B

78/92

Pep pools	Ad5-NS 10^{11} vp/dose			
	99C008	97N104	97X008	99C026
<i>F (NS3p)</i>	28	1026	579	889
<i>G (NS3h)</i>	1279	188	103	2453
<i>H (NS4)</i>	18	39	138	109
<i>I (NS5a)</i>	131	1068	172	141
<i>L (NS5b)</i>	78	144	103	32
<i>M (NS5b)</i>	24	68	47	84
<i>DMSO</i>	3	16	1	19

Pep pools	MRKAd6-NSmut 10^{11} vp/dose			
	98C047	97C055	93G	97X014
<i>F (NS3p)</i>	477	25	93	1022
<i>G (NS3h)</i>	959	398	81	1513
<i>H (NS4)</i>	36	14	99	53
<i>I (NS5a)</i>	171	45	1237	98
<i>L (NS5b)</i>	18	32	23	51
<i>M (NS5b)</i>	88	4	13	40
<i>DMSO</i>	8	3	1	5

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16C

79/92

Pep pools	MRKAd5-NSmut 10 ¹¹ vp/dose			
	99C059	99C060	97X009	96069
<i>F</i> (<i>NS3p</i>)	28	81	1308	1618
<i>G</i> (<i>NS3h</i>)	2600	161	1008	123
<i>H</i> (<i>NS4</i>)	31	74	101	40
<i>I</i> (<i>NS5a</i>)	181	99	69	96
<i>L</i> (<i>NS5b</i>)	24	31	40	20
<i>M</i> (<i>NS5b</i>)	11	58	38	164
<i>DMSO</i>	6	15	1	16

IFNy ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10¹¹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 16D

80/92

Pep pools	MRK Ad5-NSmut 10 10^10 vp/dose		
	S201	075Q	137Q
pool F (NS3p)	881	1755	73
pool G (NS3h)	573		
pool H (NS4)		3541	
pool I (NS5a)	2094		39
pool L (NS5b)			
pool M (NS5b)	756		
DMSO	319	117	44

Pep pools	MRK Ad6-NSOPTmut 10 10^10 vp/dose		
	98D209	106Q	113Q
pool F (NS3p)	5073	84	952
pool G (NS3h)	2376	160	3325
pool H (NS4)	700		
pool I (NS5a)			1106
pool L (NS5b)	530	706	
pool M (NS5b)			
DMSO	43	47	28

Pep pools	MRK Ad6-NSmut 10 10^10 vp/dose		
	S207	035Q	057Q
pool F (NS3p)	118	480	
pool G (NS3h)		196	
pool H (NS4)			
pool I (NS5a)	3340	933	
pool L (NS5b)	118		
pool M (NS5b)			
DMSO	145	34	

IFNy ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFNy/CD3/CD8 per 10^6 lymphocytes.

FIG. 17A

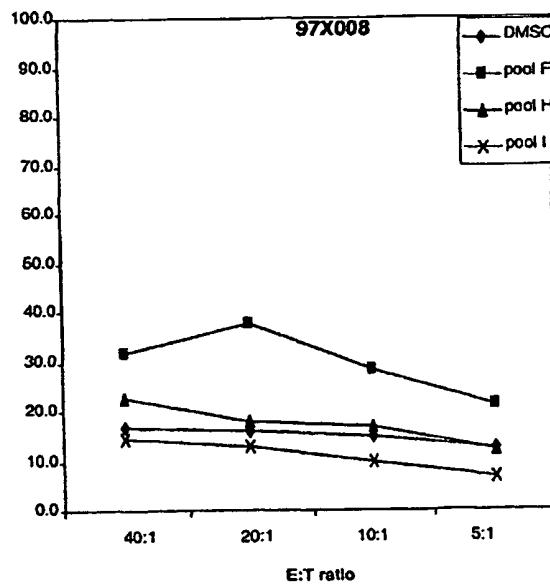
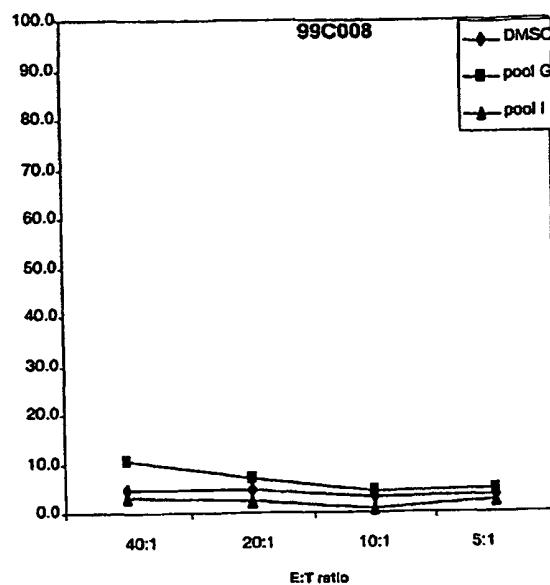
81/92

		Ad5-NS 10 ¹¹ vp/dose			
Pep pools		99C008	97N104	97X008	99C026
<i>F</i> (<i>NS3p</i>)			1703	1136	615
<i>G</i> (<i>NS3h</i>)		3153			2787
<i>H</i> (<i>NS4</i>)					
<i>I</i> (<i>NS5a</i>)			2233		
<i>L</i> (<i>NS5b</i>)					
<i>M</i> (<i>NS5b</i>)					
<i>DMSO</i>		125	98	130	0
		MRKAd6-NSmut 10 ¹¹ vp/dose			
Pep pools		98C047	97C055	93G	97X014
<i>F</i> (<i>NS3p</i>)		1024			948
<i>G</i> (<i>NS3h</i>)		3246	353		1074
<i>H</i> (<i>NS4</i>)				316	
<i>I</i> (<i>NS5a</i>)				6224	
<i>L</i> (<i>NS5b</i>)					
<i>M</i> (<i>NS5b</i>)					
<i>DMSO</i>		49	23	37	93
		MRKAd5-NSmut 10 ¹¹ vp/dose			
Pep pools		99C059	99C060	97X009	96069
<i>F</i> (<i>NS3p</i>)				2266	5053
<i>G</i> (<i>NS3h</i>)		2434	316	1018	
<i>H</i> (<i>NS4</i>)					
<i>I</i> (<i>NS5a</i>)					
<i>L</i> (<i>NS5b</i>)					
<i>M</i> (<i>NS5b</i>)				205	
<i>DMSO</i>		13	110	119	15

IFNy ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10¹¹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFNy/CD3/CD8 per 10⁶ lymphocytes.

FIG. 17B

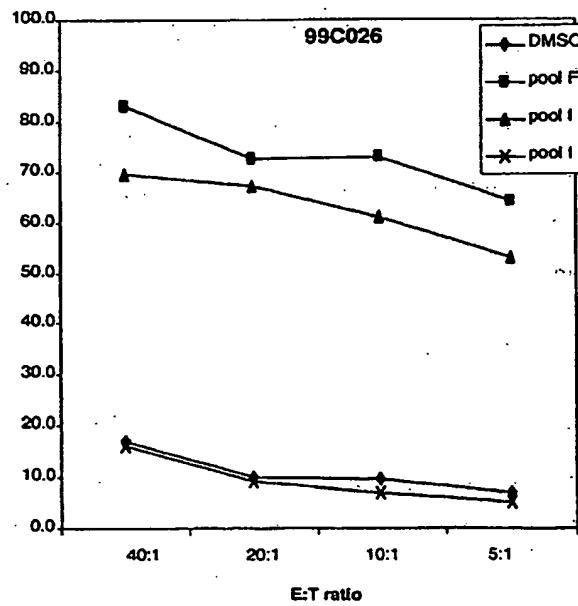
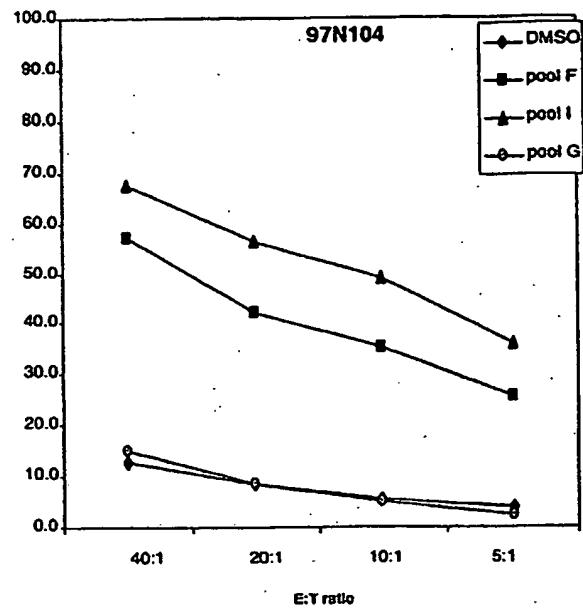
82/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Ad5-NS.

FIG. 18A

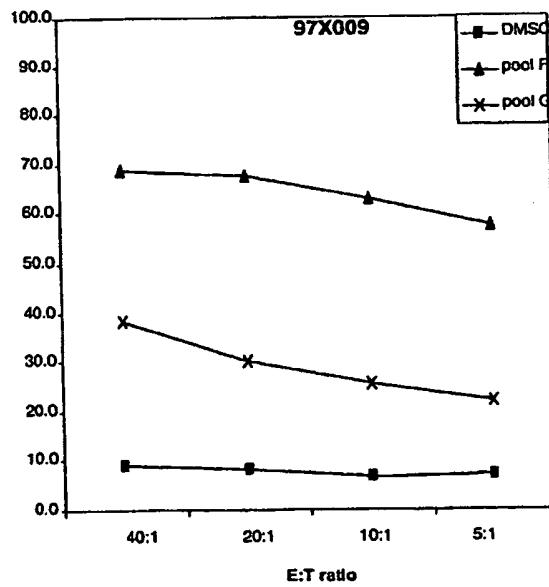
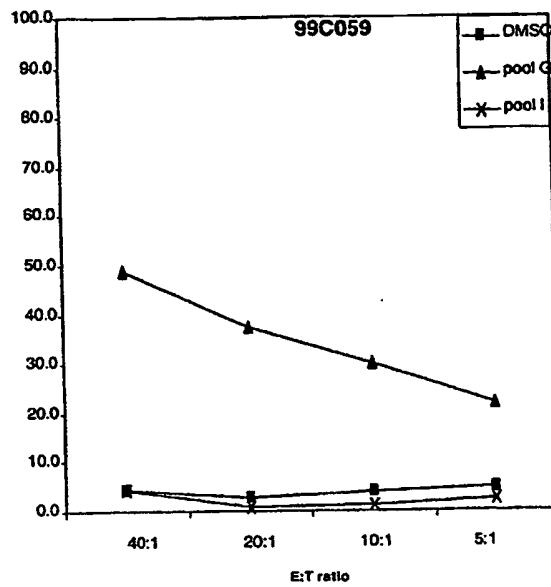
83/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Ads-NS.

FIG. 18B

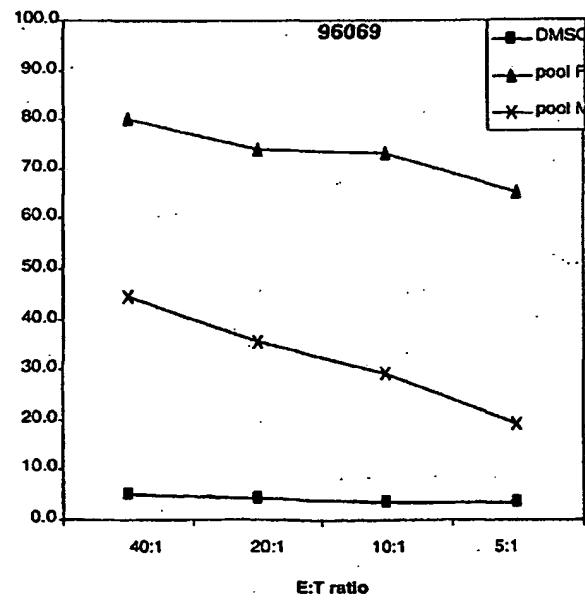
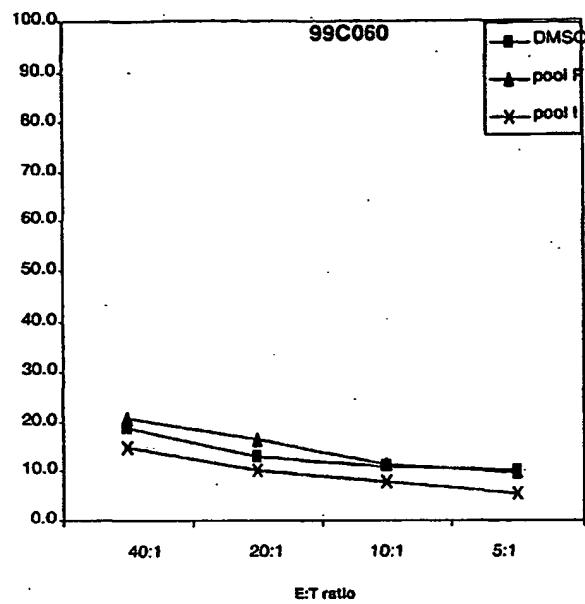
84/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd5-NSmut.

FIG. 18C

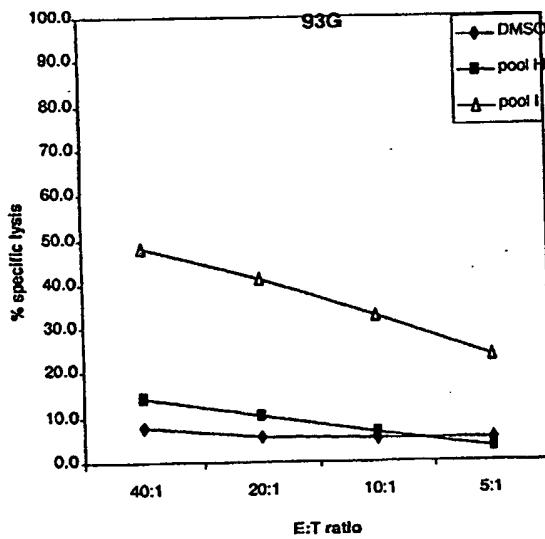
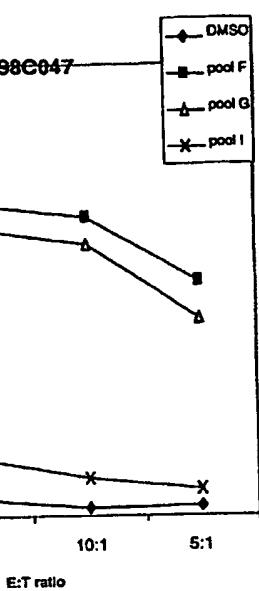
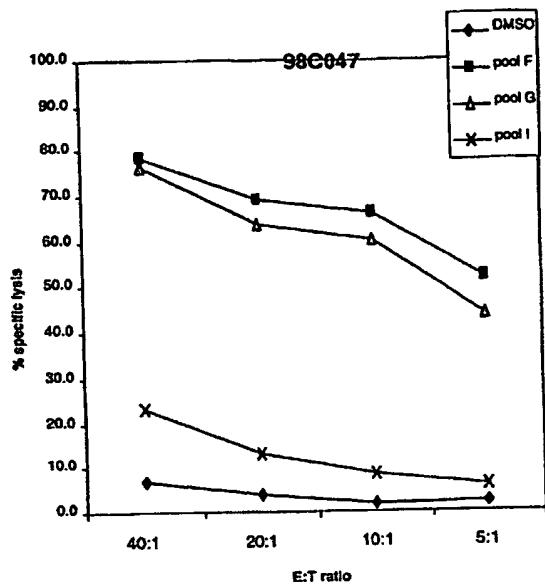
85/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd5-NSmut

FIG. 18D

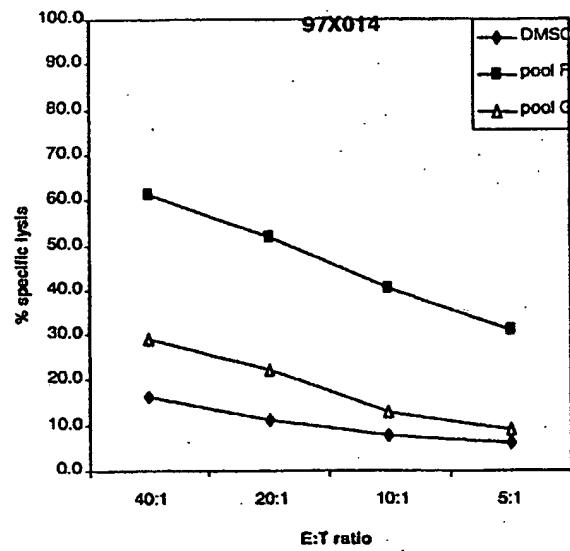
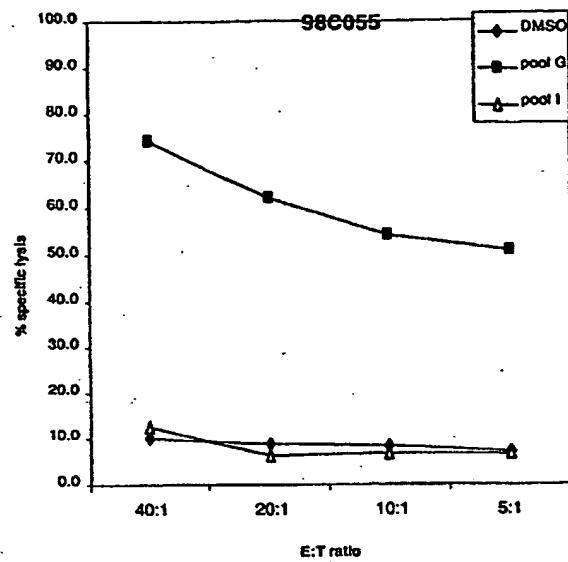
86/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd6-NSmut.

FIG. 18E

87/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd6-NSmut.

FIG. 18F

88/92

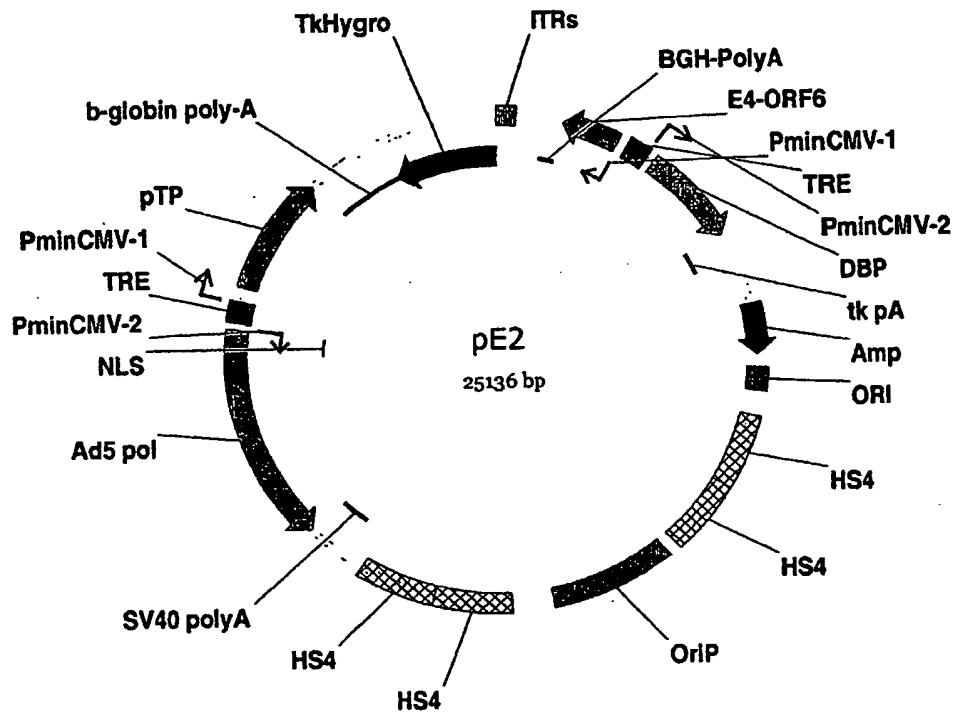


FIG. 19

89/92

1 GCCACCATGG CCCCCATCAC CGCCTACAGC CAGCAGACCA GGGGCCTGCT
51 GGGCTGCATC ATCACCGAGCC TGACCGGACG CGACAAGAAC CAGGTGGAGG
101 GAGAGGTGCA GGTGGTGAGC ACCGCTACCC AGAGCTTCCT GGCCACCTGC
151 GTGAACGGCG TGTGCTGGAC CGTGTACCAAC GGAGCCGGAA GCAAGACCCCT
201 GGCCGGACCC AAGGCCCTA TCACCCAGAT GTACACCAAT GTGGATCAGG
251 ATCTGGTGGG CTGGCAGGCC CCTCCCGGAG CCAGGAGCCT GACACCCTGT
301 ACCTGTGGAA GCAGCGACCT GTACCTGGTG ACACGCCACG CCGATGTGAT
351 CCCCGTGAGG CGCAAGGGCG ATTCTCGCGG AAGCCTGCTG AGCCCTAGGC
401 CCGTGAGCTA CCTGAAGGGC AGCAGCGGAG GACCCCTGCT GTGTCTTCT
451 GGCCATGCCG TGGGCATTTT TCGCGCTGCC GTGTGTACCA GGGCGTGGC
501 CAAAGCCGTG GATTTGTGC CCGTGGAAAG CATGGAGACC ACCATGCGCA
551 GCCCTGTGTT CACCGACAAC AGCTCTCCCC CTGCCGTGCC CCAATCATTC
601 CAGGTGGCTC ACCTGCACGC CCCTACCGGA TCTGGCAAGA GCACCAAGGT
651 GCCCGCTGCC TACGCCGCTC AGGGCTACAA GGTGCTGGTG CTGAACCCCCA
701 GCGTGGCCGC TACCTGGGC TTCGGCGCTT ACATGAGCAA GGCCCATGGC
751 ATCGACCCCCA ACATCCGCAC AGGCGTGCGC ACCATCACCA CCGGAGCTCC
801 CGTGACCTAC AGCACCTACG GCAAGTTCTT GGCGATGGA GGCTGCAGGG
851 GAGGAGCTA CGACATCATC ATCTGCGACG AGTGCACAG CACCGACAGC
901 ACCACCATCC TGGGCATTGG CACCGTGCTG GATCAGGCCG AAACAGCTGG
951 AGCCAGGCTG GTGGTGTGG CCACAGCTAC CCCTCCTGGC AGCGTGACCG
1001 TGCCCCATCC CAATATCGAG GAGGTGGCCC TGAGCAACAC AGGCGAGATC
1051 CCCTTCTACG GCAAGGCCAT CCCCCATCGAG GCCATCCGGC GAGGCAGGCA
1101 CCTGATCTTC TGCCACAGCA AGAAGAAGTG CGACGAGCTG GCTGCCAACG
1151 TGAGCGGACT GGGCATCAAC CCCGTGGCCT ACTACAGGGG CCTGGACGTG
1201 TCAGTGATCC CCACCATCGG CGATGTGGTG GTGGTGGCCA CCGACGCCCT
1251 GATGACAGGC TACACCGGAG ACTTCGACAG CGTGTACGAC TGCAACACCT
1301 GCGTGACCCA GACCGTGGAC TTCAGCCTGG ACCCCACCTT CACCATCGAA
1351 ACCACCACCG TGCCTCAGGA TGCTGTGAGC AGGAGCCAGA GGCGCGGACG
1401 CACCGGAAGG GGCAGGGCGC GAATTTATCG CTTTGTGACC CCTGGCGAAA
1451 GGCCCTCTGG CATGTTGAC AGCAGCGTGC TGTGCGAGTG CTACGACGCT
1501 GGCTGCGCTT GGTACGAGCT GACACCCGCT GAAACCAGCG TGCGCCTGCG
1551 CGCTTATCTG AATACCCCTG GCCTGCCGT GTGTCAGGAC CACCTGGAGT

FIG. 20A

90/92

1601 TCTGGGAGAG CGTGTTCACA GGACTGACCC ACATCGACGC CCATTTCTG
1651 AGCCAGACCA AGCAGGCTGG CGACAACTTC CCCTATCTGG TGGCCTATCA
1701 GGCCACCGTG TGTGCTAGGG CCCAAGCTCC ACCTCCTTCA TGGGACCAGA
1751 TGTGGAAGTG CCTGATCCGC CTGAAGCCCA CCCTGCACGG CCCTACCCCT
1801 CTGCTGTACC GCCTGGGAGC CGTGCAGAAC GAGGTGACCC TGACCCACCC
1851 CATCACCAAG TACATCATGG CCTGCATGAG CGCTGATCTG GAAGTGGTGA
1901 CCAGCACCTG GGTGCTGGTG GGAGGGTGC TGGCCGCTCT GGCTGCCTAC
1951 TGCCTGACCA CGGGAAGCGT GGTGATCGTG GGACGCATCA TCCTGAGCGG
2001 AAGGCCCCGCT ATCGTGCCCG ATCGCGAGTT CCTGTACCAAG GAGTTCGACG
2051 AGATGGAGGA GTGTGCCAGC CACCTGCCCT ACATCGAGCA GGGCATGCAG
2101 CTGGCCGAAC AGTTCAAGCA GAAGGCCCTG GGCCTGCTGC AGACAGCCAC
2151 CAAACAGGCC GAAGCTGCCG CTCCCGTGGT GGAAAGCAAG TGGAGGGCCC
2201 TGGAGACCTT CTGGGCTAAG CACATGTGGA ACTTCATCTC TGGCATCCAG
2251 TACCTGGCCG GACTGAGCAC CCTGCCTGGC AACCCCGCTA TCGCCAGCCT
2301 GATGGCCTTC ACCGCTAGCA TCACCTCTCC CCTGACCACC CAGAGCACCC
2351 TGCTGTTCAA CATTCTGGGC GGATGGGTGG CCGCTCAGCT GGCCCCCTCCT
2401 TCAGCTGCTT CTGCCCTTGT GGGCGCTGGC ATTGCCGGAG CCGCTGTGGG
2451 CAGCATTGGC CTGGCAAAG TGCTGGTGG A TATTCTGGCT GGCTATGGCG
2501 CTGGCGTGGC CGGAGCCCTG GTGGCCTTCA AGGTGATGAG CGGAGAGATG
2551 CCCAGCACCG AGGACCTGGT GAACCTGCTG CCTGCCATTG TGAGCCCTGG
2601 AGCCCTGGTG GTGGCGTGG TGTGTGCTGC CATTCTGAGG CGCCATGTGG
2651 GACCCGGAGA GGGCGCTGTG CAGTGGATGA ACCGCCTGAT CGCCTTCGCC
2701 TCTCGCGGAA ACCACGTGAG CCCTACCCAC TACGTGCCCTG AGAGCGACGC
2751 CGCTGCCAGG GTGACCCAGA TCCTGAGCAG CCTGACCATC ACCCAGCTGC
2801 TGAAGGCCT GCACCAGTGG ATCAACGAGG ACTGCAGCAC ACCCTGCAGC
2851 GGAAGCTGGC TGAGGGACGT GTGGGACTGG ATCTGCACCG TGCTGACCGA
2901 CCTCAAGACC TGGCTGCAGA GCAAGCTGCT GCCCCAACTG CCTGGCGTGC
2951 CCTTCTTCTC ATGCCAGCGC GGATACAAGG GCGTGTGGAG GGGCGATGGC
3001 ATCATGCAGA CCACCTGTCC CTGGCGAGCC CAGATCACAG GCCACGTGAA
3051 GAACGGCAGC ATGCCATCG TGGGCCCTAA GACCTGCAGC AACACCTGGC
3101 ACGGCACCTT CCCCATCAAC GCCTACACCA CCGGACCCCTG CACACCCAGC
3151 CCTGCTCCCA ACTACAGCAG GGCCCTGTGG AGGGTGGCTG CCGAGGAGTA

FIG. 20B

91/92

3201 CGTGGAGGTG ACCAGGGTGG GAGACTTCCA CTACGTGACC GGAATGACCA
3251 CCGACAAACGT GAAGTGTCCC TGTCAAGGTGC CCGCTCCCGA ATTTTTTACCC
3301 GAAAGTGGATG GCGTGCAGCT GCATCGCTAT GCCCCCTGCCT GTAGGGCCCT
3351 GCTGCGCGAA GAAGTGACCT TCCAGGTGGG CCTGAACCAAG TACCTGGTGG
3401 GCAGCCAGCT GCCCTGCGAG CCTGAGCCCG ATGTGGCCGT GCTGACCCAGC
3451 ATGCTGACCG ACCCCAGCCA CATCACAGCC GAAACCGCTA AAAGGCGCCT
3501 GGCCAGGGGC TCTCCTCCAA GCCTGGCCCTC AAGCAGCGCT AGCCAGCTGT
3551 CTGCTCCCAG CCTGAAGGCC ACCTGACCCA CCCACCACGT GAGCCCCGAC
3601 GCCGACCTGA TCGAGGCCAA CCTGCTGTGG CGCCAGGAGA TGGGCGGCAA
3651 CATCACCCGC GTGGAGAGCG AGAACAAAGGT GGTGGTGCTG GACAGCTTCG
3701 ACCCCCTGCG CGCCGAGGAG GACGAGCGCG AGGTGAGCGT GCCCCGCCAG
3751 ATCCTGCGCA AGAGCAAGAA GTTCCCCGCT GCCATGCCCA TCTGGGCTAG
3801 ACCTGATTAC AACCCCTCCCC TGTGGAGAG CTGGAAGGAC CCTGATTACG
3851 TGCCCTCCAGT GGTGCATGGC TGTCCCTCTGC CTCCCATTAAGGCCCCCTCCT
3901 ATTCCACCTC CTAGGCAGCA AAGGACCGTG GTGCTGACAG AAAGCAGCGT
3951 GAGCTCTGCT CTGGCCGAAC TGGCCACCAA GACCTTGCG AGCAGCGAGA
4001 GCTCTGCCGT GGACAGCGGA ACAGCCACCG CTCTGCCCTGA CCAGGCCAGC
4051 GACGACGGCG ATAAGGGCAG CGATGTGGAG AGCTATAGCA GCATGCCCTCC
4101 CCTGGAAGGC GAACCTGGCG ATCCCGATCT GAGCGATGGC AGCTGGAGCA
4151 CCGTGAGCGA AGAGGCCAGC GAGGACGTGG TGTGTTGCAG CATGAGCTAC
4201 ACCCTGGACAG GCGCTCTGAT CACACCCCTGC GCTGCCGAGG AGAGCAAGCT
4251 GCCCCATCAAC GCCCTGAGCA ACAGCCTGCT GAGGCACCAAC AACATGGTGT
4301 ACGCCACCCAC CAGCAGGTCT GCCGGACTGA GGCAGAAGAA GGTGACCTTC
4351 GACCGCCCTGC AGGTGCTGGA CGACCACTAC CGCGATGTGC TGAAGGAGAT
4401 GAAGGCCAAG GCCAGCACCG TGAAGGCCAA GCTGCTGAGC GTGGAGGAGG
4451 CCTGCAAGCT GACCCCCCCCC CACAGCGCCA AGAGCAAGTT CGGCTACGGC
4501 GCCAAGGACG TGGCGAACCT GAGCAGCAAG GCCGTGAACC ACATCCACAG
4551 CGTGTGGAAG GACCTGCTGG AGGACACCGT GACCCCCATC GACACCCACCA
4601 TCATGGCCAA GAACGAGGTG TTCTGCGTGC AGCCCGAGAA GGGCGGCCGC
4651 AAGCCCGCTC GCCTGATCGT GTTCCCCGAT CTGGGCGTGC GCGTGTGCGA
4701 GAAGATGGCC CTGTACGACG TGGTGAGCAC CCTGCCTCAG GTGGTGTGATGG
4751 GCTCAAGCTA CGGCTTCCAG TACAGCCCTG GCCAGCGCGT GGAGTTCCCTG

FIG. 20C

92/92

4801 GTGAACACCT GGAAGAGCAA GAAGAACCCC ATGGGCTTCA GCTACGACAC
4851 ACGCTGCTTC GACAGCACCG TGACCGAGAA CGACATCCGC GTGGAGGAGA
4901 GCATCTACCA GTGCTGCGAC CTGGCCCTG AGGCCAGGCA GGCCATCAAG
4951 AGCCTGACCG AGCGCTGTA CATCGGAGGC CCTCTGACCA ACAGCAAGGG
5001 ACAGAACTGC GGATACAGGC GCTGTAGGGC CTCTGGCGTG CTGACCACCA
5051 GCTGTGGCAA CACCCCTGACC TGCTACCTGA AGGCCAGCGC TGCCCTGTCGC
5101 GCTGCCAACG TGCAAGACTG CACCATGCTG GTGAACGCCG CTGGCCTGGT
5151 GGTGATTTGT GAAAGCGCTG GCACCCAGGA AGATGCTGCC AGCCTGCGCG
5201 TGTTCACCGA GGCCATGACC AGGTACTCTG CCCCTCCCGG AGACCCCCCT
5251 CAGCCCCAAT ACGACCTGGA GCTGATCACC AGCTGCTCAA GCAACGTGAG
5301 CGTGGCTCAC GACGCCAGCG GAAAGCGCGT GTACTACCTG ACACCGCATC
5351 CCACCACCCC TCTGGCTCGC GCTGCCCTGG AAACCGCTCG CCATACACCC
5401 GTGAACAGCT GGCTGGGCAA CATCATCATG TACGCCCTA CCCTGTGGGC
5451 TCGCATGATC CTGATGACCC ACTTCTTCAG CATCCTGCTG GCTCAGGAGC
5501 AGCTGGAGAA GGCCCTGGAC TGCCAGATT ACAGGCCTTG CTACAGCATC
5551 GAGCCCCCTGG ACCTGCCCA AATCATCGAG CGCCTGACG GCCTGTCTGC
5601 CTTCAGCCTG CACAGCTACA GCCCTGGCGA AATTAATCGC GTGGCCAGCT
5651 GTCTGCGCAA ACTGGCGTG CCTCCTCTGC GCGTGTGGAG GCATAGGGCT
5701 AGGAGCGTGA GGGCTAGGCT GCTGAGCCAG GGAGGCAGGG CCGCTACCTG
5751 TGGAAAGTAC CTGTTCAACT GGGCCGTGAA GACCAAGCTG AAGCTGACCC
5801 CTATCCCTGC CGCTAGCCAG CTGGACCTGA GCGGATGGTT CGTGGCTGGC
5851 TACAGCGGAG GCGACATCTA CCACAGCCTG TCTCGCGCTC GCCCTCGCTG
5901 GTTCATGCTG TGCCCTGCTGC TGCTGAGCGT GGGCGTGGGC ATCTACCTGC
5951 TGCCCCAACCG CTAAA

FIG. 20D

IN THE PCT RECEIVING OFFICE
OF THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	Merck & Co., Inc		
PCT Serial No.:	To Be Assigned	Case No.:	PCT ITR0015Y
Filing date:	On Even Date Herewith		
For:	HEPATITIS C VIRUS VACCINE		
		US/RO	
		Authorized Officer:	To Be Assigned

Assistant Commissioner of Patents
BOX PCT
Washington, D.C. 20231

**NUCLEOTIDE AND/OR AMINO ACID
SEQUENCE DISCLOSURE, PCT RULE 5.2**

Sir:

As required under PCT Rule 5.2, Applicant respectfully encloses a paper (64 pages) and a computer readable form of the Sequence Listing for the above-identified PCT International Application, filed on even date herewith.

I hereby state that the content of the paper and computer readable forms of the Sequence Listing, submitted in accordance with WIPO and Standard ST.23 and under PCT Rule 13ter.1, respectively, are the same.

Respectfully submitted,

By Sheldon Heber
Sheldon O. Heber
Reg. No. 38,179
Attorney for Applicants

Merck & Co., Inc.
P.O. Box 2000
Rahway, NJ 07065-0907
(732) 594-1958

SEQUENCE LISTING

<110> Merck & Co. Inc., and Istituto Di Ricerche Di Biologia Molecolare P. Angeletti S.P.A.

<120> HEPATITIS C VIRUS VACCINE

<130> ITR0015Y

<150> 60/363,774

<151> 2002-03-13

<150> 60/328,655

<151> 2001-10-11

<160> 17

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1985

<212> PRT

<213> Artificial Sequence

<220>

<223> Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide

<400> 1

Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
1 5 10 15
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
20 25 30
Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys
35 40 45
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr
50 55 60
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp
65 70 75 80
Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr
85 90 95
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
100 105 110
Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
115 120 125
Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
130 135 140
Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys
145 150 155 160
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met
165 170 175
Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro
180 185 190
Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly
195 200 205

Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr
 210 215 220
 Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly
 225 230 235 240
 Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
 245 250 255
 Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly
 260 265 270
 Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
 275 280 285
 Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile
 290 295 300
 Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
 305 310 315 320
 Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn
 325 330 335
 Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly
 340 345 350
 Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe
 355 360 365
 Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly
 370 375 380
 Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val
 385 390 395 400
 Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met
 405 410 415
 Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys
 420 425 430
 Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
 435 440 445
 Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly
 450 455 460
 Arg Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly
 465 470 475 480
 Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
 485 490 495
 Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val
 500 505 510
 Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
 515 520 525
 His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp
 530 535 540
 Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr
 545 550 555 560
 Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
 565 570 575
 Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
 580 585 590
 Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Gln Asn
 595 600 605
 Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met
 610 615 620
 Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly
 625 630 635 640

Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val
 645 650 655
 Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp
 660 665 670
 Arg Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser
 675 680 685
 His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys
 690 695 700
 Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala
 705 710 715 720
 Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp
 725 730 735
 Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly
 740 745 750
 Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe
 755 760 765
 Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe
 770 775 780
 Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala
 785 790 795 800
 Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser
 805 810 815
 Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala
 820 825 830
 Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met
 835 840 845
 Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro
 850 855 860
 Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His
 865 870 875 880
 Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala
 885 890 895
 Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu
 900 905 910
 Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile
 915 920 925
 Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser
 930 935 940
 Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys
 945 950 955 960
 Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro
 965 970 975
 Gln Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly
 980 985 990
 Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala
 995 1000 1005
 Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro
 1010 1015 1020
 Lys Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr
 1025 1030 1035 1040
 Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala
 1045 1050 1055
 Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly
 1060 1065 1070

Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro
1075 1080 1085
Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg
1090 1095 1100
Leu His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val
1105 1110 1115 1120
Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro
1125 1130 1135
Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp
1140 1145 1150
Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly
1155 1160 1165
Ser Pro Pro Ser Leu Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro
1170 1175 1180
Ser Leu Lys Ala Thr Cys Thr His His Val Ser Pro Asp Ala Asp
1185 1190 1195 1200
Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile
1205 1210 1215
Thr Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp
1220 1225 1230
Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu
1235 1240 1245
Ile Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala
1250 1255 1260
Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp
1265 1270 1275 1280
Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala
1285 1290 1295
Pro Pro Ile Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu
1300 1305 1310
Ser Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly
1315 1320 1325
Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro
1330 1335 1340
Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr
1345 1350 1355 1360
Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser
1365 1370 1375
Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val
1380 1385 1390
Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys
1395 1400 1405
Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu
1410 1415 1420
Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly
1425 1430 1435 1440
Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp
1445 1450 1455
His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val
1460 1465 1470
Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro
1475 1480 1485
His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn
1490 1495 1500

Leu Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu
 1505 1510 1515 1520
 Leu Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn
 1525 1530 1535
 Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg
 1540 1545 1550
 Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala
 1555 1560 1565
 Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser
 1570 1575 1580
 Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn
 1585 1590 1595 1600
 Thr Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg
 1605 1610 1615
 Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser
 1620 1625 1630
 Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys
 1635 1640 1645
 Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys
 1650 1655 1660
 Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr
 1665 1670 1675 1680
 Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala
 1685 1690 1695
 Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Ala Ala
 1700 1705 1710
 Gly Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala
 1715 1720 1725
 Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro
 1730 1735 1740
 Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys
 1745 1750 1755 1760
 Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr
 1765 1770 1775
 Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu
 1780 1785 1790
 Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met
 1795 1800 1805
 Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe
 1810 1815 1820
 Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln
 1825 1830 1835 1840
 Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile
 1845 1850 1855
 Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser
 1860 1865 1870
 Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val
 1875 1880 1885
 Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg
 1890 1895 1900
 Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe
 1905 1910 1915 1920
 Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala
 1925 1930 1935

Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly
 1940 1945 1950
 Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu
 1955 1960 1965
 Cys Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn
 1970 1975 1980

Arg
 1985

<210> 2
 <211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> Non-optimized cDNA sequence encoding SEQ. ID. NO.
1

<400> 2
 gccaccatgg cgcccatcac ggcctactcc caacagacgc gggccctact tggttgcac
 atcactagcc ttacaggccg ggacaagaac caggtcgagg gagaggttca ggtggttcc
 accgcacac aatcccttc ggcgacactgc gtcacacggcg tgggttggac cgttacat
 ggtgctggct caaagacctt agccggccca aaggggccaa tcacccagat gtacactaat
 gtggaccagg acctcgctgg ctggcaggcg ccccccgggg cgcgttccctt gacaccatgc
 acctgtggca gtcagacct ttacttggc acgagacatg ctgacgtcat tccgggtgc
 cggcggggcg acagtagggg gaggctgtc tccccaggc ctgtctcta cttgaaggc
 tcttcgggtg gtcactgtc ctgccttcg gggcagctg tgggcattt cccggctg
 gtatgcaccc ggggggttgc gaaggcggtg gactttgtgc ccttagatc catggaaact
 actatgcgt ctccggtctt cacggacaac tcatcccccc cggccgtacc gcagtcat
 caagtggccc acctacacgc tcccactggc agcggcaaga gtactaaagt gccggctg
 tatgcagccc aagggtacaa ggtgctgtc ctcaatccgt ccgttggc taccttaggg
 tttggggcgt atatgtctaa ggcacacggt attgacccca acatcagaac tggggtaagg
 accattacca caggcgcccc cgtcacatac tctacatcg gcaagtttct tgccgatgt
 ggttgcctg gggcgctta tgacatcata atatgtatg agtgcattt aactgactcg
 actacaatct tggcactgg cacagtcgt gacaaaggcg agacggctgg agcggcgtt
 gtcgtgtcg ccaccgtac gcctccggg tgggtcaccg tgccacaccc aaacatcgag
 gaggtggccc tgcataatac tggagagatc cccttctatg gcaaagccat cccattgaa
 gccatcaggg ggggaaggca tctcatttc tgcattcca agaagaagtgc gacgcgc
 gcccaaagc tgcaggccctt cggaaatcaac gctgtggcgt attaccggg gtcgatgt
 tccgtcatac caactatcg agacgtcggt gtcgtggcaa cagacgtct gatgacggc
 tatacggcg actttgactt agtgcac gtaacacat gtcgcatttca gacagtgcac
 ttcagcttg atcccaccc taccatttag agacgaccg tgcctcaaga cgcagtgt
 cgctcgac ggcggggtag gactggcagg gtaggagag gcatctacag gttgtact
 cggggagaac ggcctcggtt catgtcgat tccctcggtcc tgggtgatg ctatgcgc
 ggctgtgctt ggtacgagct caccggccg gagacctcggtt taggtgcg ggcctacctg
 aacacaccag ggttgcctgttgcaggac cacctggagt tctggagag tgcatttca
 ggcctaccc acatagatgc acacttcttgc tcccaacca agcaggcagg agacaactt
 ccctacccgtt tagcatacca agccacgggt tgcggcagggtt ctcaggcccc acctccatca
 tggatcaa tggatggatg ttcataccgt ctgaaaccta cgctgcacgg gccaacaccc
 ttgcgttaca ggctggggagg cgtccaaat gaggtcacccttcaacccccc cataaccaaa
 tacatcatgg catgcgttc ggcgtaccc tgcgtaccc tgcgtaccc tgcgttgc
 ggcggagtcc ttgcgtctt ggcgggtt ggcgttat tgcctgacaa caggcagtgt
 ggtggatgatg ttcgttccgg gaggccggat attgttcccg acagggagtt tctctaccag
 gagttcgatg aaatggaa gtcgcctcg cacctccctt acatcgacca gggatgc
 ctgcggcggc aattcaagca gaaagcgctc gggtaactgc aaacagccac caaacaagcg

gaggctgctg	ctcccggtgt	ggagtccaaag	ttggcagcccc	tttagacatt	ctggggcaag	2220
cacatgtgga	atttcatcag	cgggatacag	tacttagcag	gtttatccac	tctgcctggg	2280
aaccccgcaa	tagcatcatt	gatggcattc	acagcctcta	tcaccagccc	getcaccacc	2340
caaagtaccc	tcctgtttaa	catcttgggg	gggtgggtgg	ctgcccact	cgcccccccc	2400
agcgcgcctt	cggtttcggt	gggcgcggc	atcgccgtg	cggctgttgg	cagcataggc	2460
cttgggaagg	tgcttgcga	cattctggcg	ggttatggag	caggagtggc	cggcgcgc	2520
gtggccttca	aggtcatgag	cggcgagatg	ccctccaccc	aggacctgtt	caatctactt	2580
cctgcacatcc	tctctcctgg	cgccctggc	gtcggttgc	tgtgtgcagc	aatactgcgt	2640
cgacacgtgg	gtccgggaga	gggggctgtg	cagttgtatg	accggctgtat	agcgttcgc	2700
tcgcggggtt	atcatgttcc	ccccacgcac	tatgtgcctg	agagcgcacgc	cgcagcgcgt	2760
gttactcaga	tcctctccag	ccttaccatc	actcagctgc	tggggatgt	ttgggactgg	2820
attaatgcagg	actgtccac	accgtgttcc	ggctgtggc	ttgggactgg	2880	
atatgcacgg	tgttgcactga	cttcaagacc	tggctccagt	ccaagctct	gcccgcacgt	2940
ccgggatcc	ctttttctc	gtgccaacgc	gggtacaagg	gagtctggc	gggagacggc	3000
atcatgaaa	ccacctgccc	atgtggagca	cagatcaccg	gacatgtcaa	aaacggttcc	3060
atgaggatcg	tcggggctaa	gacctgcagc	aacacgttgc	atggAACATT	ccccatcaac	3120
gcatacacca	cggggccctg	cacaccctt	ccagcgcacaa	actattctag	ggcgcgtgtgg	3180
cgggtggccg	ctgaggagta	cgtggaggc	acgcgggtgg	gggatttcca	ctacgtgacg	3240
ggcatgacca	ctgacaacgt	aaagtgcaca	tgcagggtt	cggctccctg	atttctcaccg	3300
gagggtggacg	gagtgcgggtt	gcacaggta	gctccgggt	gcaggcctt	cctacgggag	3360
gagggttacat	tccagggtcgg	gctcaaccaa	tacctgggtt	ggtcacagat	accatgcgag	3420
cccgaaacgg	atgttagcgt	gctacttcc	atgtcaccgg	acccctccca	catcacagca	3480
gaaacggcta	agcgttaggtt	ggccaggggg	tctcccccct	ccttggccag	ctcttcagct	3540
agccagggtt	ctgcgccttc	cttgaaggcg	acatgcacta	cccaccatgt	ctctccggac	3600
gctgacacta	tcgaggccaa	cctcctgtgg	cgccaggaga	tggggggaa	catcacccgc	3660
gtggagtcgg	agaacaaagg	ggtagtcctg	gactcttgc	acccgcttcg	agcggaggag	3720
gatgagaggg	aagtatccgt	tccggggag	atctgcgg	aatccaagaa	gttccccgca	3780
gcgatgcaca	tctgggcccgg	cccggattac	aaccctccac	tgttagagtc	ctggaaaggac	3840
ccggactacg	tccctccgg	ggtgcacggg	tgcccggttgc	cacctatcaa	ggcccccctca	3900
ataccacctc	cacggagaaa	gaggacgggtt	gtctcaacag	agtccctcg	gtcttcgtcc	3960
ttagcggagc	tcgtactaa	gaccttcggc	agtcctcgaa	catcgccgt	cgacagcggc	4020
acggcggacgg	cccttcctg	ccaggcctcc	gacgcgggtg	acaaaggatc	cgacgttgag	4080
tcgtactct	ccatggcccc	ccttgggggg	gaaccggggg	accccgatct	cagtacggg	4140
tcttgggtcta	ccgtgagcg	ggaagcttagt	gaggatgtcg	tctgctgtc	aatgtccctac	4200
acatggacag	gcccgttgc	cacgcctatgc	gctgcggagg	aaagcaagct	gcccatcaac	4260
gcgttgcac	actcttcgt	gcccacccat	aaatggttt	atgccacaac	atctcgcage	4320
gcaggcctgc	ggcagaagaa	ggtcacctt	gacagactgc	aagtccctg	cgaccactac	4380
cgggacgtgc	tcaaggagat	gaaggcgaag	gcgtccacag	ttaaggctaa	actccatcc	4440
gtagaggaag	cctgcacgat	gacgccccca	cattcggcca	aatccaagtt	tggctatggg	4500
gcaaaggacg	tccggaaacct	atccagcaag	gcofttaacc	acatccactc	cgtgtggaa	4560
gactgtgtgg	aagacactgt	gacaccaatt	gacaccacca	tcatggcaaa	aatagagggtt	4620
ttctgtgtcc	aaccagagaaa	aggaggccgt	aagccagccc	gccttatacg	attcccgat	4680
ctgggagtcc	gtgtatgcga	gaagatggcc	ctctatgt	tgttctccac	ccttccctcg	4740
gtctgtatgg	gctccctata	cggattcccg	tactctctg	ggcagcgt	cgagttcccg	4800
gtgaataacct	gaaaatcaaa	aaaaaacc	atgggtttt	catatgacac	tcgctgtttc	4860
gactcaacgg	tcacccggaaa	cgacatccgt	gttgaggagt	caatttacca	atgtgtgac	4920
ttggggcccg	aagccagaca	ggccataaaa	tcgctcacag	agcggctta	tatccccgg	4980
cctctgacta	attcaaaaagg	gcagaactgc	ggttatacgcc	ggtgccgc	gagccggcgt	5040
ctgacgacta	gtcgccgtaa	caccctcaca	tgttacttga	aggcctctgc	agcctgtcga	5100
gctgcaagc	tccaggactg	cacgatgtc	gtgaacgcgc	ccggccttgc	cgttatctgt	5160
gaaagcgcgg	gaacccaaga	ggacgcggcg	agcctacgc	tcttcacgg	ggctatgact	5220
aggtactctg	ccccccccccgg	ggacccccc	caaccagaat	acgacttgg	gtctgataaca	5280
tcatgttcc	ccaatgtgtc	ggtgcggccac	gtgcacatcg	gcaaaaagggt	gtactacctc	5340
acccgtgate	ccaccacccc	cctcgcacgg	gctgcgtggg	aaacagctag	acacactcca	5400
gttaactcc	ggcttaggcaa	cattatcatg	tatgcgcacca	ctttgtggc	aaggatgatt	5460

ctgatgactc acttcttctc	catccttcta gcacaggagc	aacttggaaaa agccctggac	5520
tgccagatct acggggcctg	ttactccatt gagccacttg	acctaccta gatcattgaa	5580
cgactccatg gccttagcgc	atttcactc catagttact	ctccaggtga gatcaatagg	5640
gtggcttcat gcctcaggaa	acttgggta ccacccttgc	gagtctggag acatcgggccc	5700
aggagcggtcc ggcgttaggt	actgtcccgaggggg	ccgcccacttg tggcaagttac	5760
ctcttcaact gggcagtggaa	gaccaaactc aaactcactc	caatcccggc tgcgtcccag	5820
ctggacttgtt cggctgggtt	cgttgctgtt tacaggggg	gagacatata tcacagctg	5880
tctctgtccc gaccccgctg	gttcatgtt tgcttactcc	tactttctgtt agggtaggc	5940
atctacccatgc tcccaaccg	ataaa		5965

<210> 3

<211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> Optimized cDNA encoding SEQ ID NO: 1

<400> 3

gccaccatgg ccccatcac	cgccctacagc cagcagaccc	gcggcctgct gggctgcata	60
atcaccagcc tgaccggccg	cgacaagaac cagggtggagg	gcgagggtca ggtgggtgagc	120
accggccaccc agagcttctt	ggccacatgc gtgaacggcg	tgtgctgac cgtgtaccac	180
ggcgccggca gcaagaccc	ggccggccccc aaggggccca	tcacccagat gtacaccaac	240
gtggaccagg acctgggtgg	ctggcaggcc cccccccggc	cccgccagct gaccggcgc	300
acettggccgca gcacgcact	gtacccctgtt acccccccac	ccgacgtat cccctgtgc	360
cgccggccgac acagccggg	caggctgtt agccccccgc	ccgtgagctt cctgaagggc	420
agcagccggcg gccccctgtt	gtgccccagc ggcacgccc	tggcatctt cccgcggcc	480
gtgtgcaccc gggcggtggc	caaggccgtt gacttcgtgc	ccgtggagag catggagacc	540
accatgcgcac gccccgtgtt	caccgacaaac agcagcccc	ccggccgtgc ccagagctt	600
cagggtggccc acctgcacgc	ccccacccgc agccggcaaga	gcaccaaggt gccccggcc	660
tacggccccc agggctacaa	gttgctgggtt ctgaacccca	gcgtggccgc caccctgggc	720
ttcggccct acatgagcaa	ggccccacggc atcgacccca	acatccgcac cggcggtgc	780
accatcacca cccggccccc	cgtgacccat acgacaccc	gcaagttctt gggccacggc	840
ggctgcacgc gggccgccta	cgacatcatac atctgcgac	agtgcacag caccgacgc	900
accaccatcc tggcatacg	caccgtgtt gaccaggccg	agaccgcgg cggccgcctg	960
gtgggtgtgg ccacccac	cccccccgcc agcgtgaccc	tggcccccacca caacatcgag	1020
gagggtggccc tgaccaac	cgccgagatc ccccttctacg	gcaaggccat cccatcgag	1080
gccatccgcg gggccgcac	cctgatctt tgcacacagca	agaagaatg cgacgagctt	1140
ggcccaacgc tgacggccct	ggccatcaac gcccgtggct	actaccgcgg cttggacgtt	1200
agcgtgatcc ccaccatcg	cgacgtgggtt gtgtggcca	ccgacgcctt gatgacccgc	1260
tacacccgcg acttcgacag	cgatcgatcgac tgcacacac	gcgtgaccca gaccgtggac	1320
ttcagcctgg accccaccc	caccatcgat accaccaccc	tggcccaaggc cgccgtgagc	1380
cgccagccgc gccggccgc	caccggccgc ggccggccgc	gcatctaccg cttcgtgacc	1440
cccgccgagc gccccagcgg	catgttcgac agcagctgc	tgtgcaatgt ctacgaccc	1500
ggctgcgcct ggtacgac	gaccccccggc gagaccagcg	tgcgcctgcg cgccctacgt	1560
aacacccccc gctgtccct	gtgcccaggac caccctggat	tctggagag cgttttacc	1620
ggcctgaccc acatcgacgc	ccacttctt agccagacca	agcaggccgg cgacaactt	1680
ccctacttgg tggcatacca	ggccacccgtt tgcccccgc	cccaggcccc cccccccgc	1740
tgggaccaga tggaaatgt	cctgatccgc ctgaagccca	ccctgcacgg cccacccccc	1800
ctgctgtacc gcctggccgc	cgtgcacaaac gaggtggacc	tgacccaccc cataccaag	1860
tacatcatgg cctgatcg	cgccgacatcg gaggtggat	ccagcacatg ggtgttgg	1920
ggccggcgtgc tgccgcctt	ggccgcctac tgccgtacca	ccggcagctt ggtgttgg	1980
ggccgcatac tccctgagc	ccgcccccc atcgcccc	accgcgatgt cctgtaccag	2040
gagttcgacg agatggagga	gtgcgcacgc caccctggct	acatcgacca gggcatgcag	2100
ctggccgagc agttcaagca	gaaggccctg ggccgtgtc	agaccgcac caaggccgc	2160

gaggccgccc	cccccggt	ggagagcaag	tggcgcccc	tggagacctt	ctgggccaag	2220
cacatgtgga	acttcatcg	cgccatccag	tacctggccg	gcctgagcac	cctgccccgc	2280
aaccggcca	tcggcagct	gatggcccttc	accggccagca	tcaccagccc	cctgaccacc	2340
cagagcaccc	tgtgttcaa	catccctgggc	ggctgggtgg	ccggcccagct	ggcccccccc	2400
agcgccgcca	gccccttcgt	gggcgccccg	atcgccggcg	ccggcgtggg	cagcatcgcc	2460
ctgggcaagg	tgtgttggga	catccctggcc	ggctacggcg	ccggcgtggc	ccggccccctg	2520
gtggcctca	aggtgtatgag	cggegagatg	cccagcaccc	aggacctgtt	gaacctgtcg	2580
cccgccatcc	tgagccccgg	cgccctgggt	gtgggctgtt	tgtgcgccc	catcctgcgc	2640
cgccacgtgg	gccccggcg	gggcgccccgt	cagtggatga	accgcctgtat	cgccttcgccc	2700
agccggcca	accacgttag	ccccaccac	tacgtgcccc	agagcgcacgc	cgccggccccgc	2760
gtgaccaga	tccctgagcag	cctgaccatc	accaggctgc	tgaagcgcct	gcaccaggctg	2820
atcaacgagg	actgcagcac	ccctgcage	ggcagctggc	tgccgcgacgt	gtgggactgg	2880
atctgcaccg	tgtgtgaccga	cttcaagacc	tggctgcaga	gcaagctgt	gccccagctg	2940
cccgccgtgc	ccttcttcag	ctgcccgc	ggctacaagg	gcgtgtggcg	cggcgcacggc	3000
atcatgcaga	ccacccgtccc	ctggccgc	catatcaccg	gcccacgtgaa	gaacggcagc	3060
atgcgcatcg	tggggcccca	gacctgcagc	aacacctggc	acggcacctt	ccccatcaac	3120
gcctacacca	ccggccccctg	caccccccac	cccgccccca	actacagccg	cgccctgtgg	3180
cgcgccggcg	cccgaggagta	cgtggaggtg	acccgcgtgg	gcgacttcca	ctacgtgacc	3240
ggcatgacca	ccgacaacgt	gaagtggccc	tgcagggtgc	ccgcggggcg	gttcttcacc	3300
gagggtggacg	gcgtgcgcct	gcaccgtac	gccccccgcct	gcccggggcg	gtgcgcgag	3360
gagggtgacct	tccagggtgg	cctgaaccag	tacccgttgg	gcagccagct	gcccctgcgag	3420
cccgagcccc	acgtggccgt	gctgaccaggc	atgtgtaccg	accccaagcca	catcaccggcc	3480
gagaccggcca	agcgccccct	ggccggccgc	agccccccca	gcctggccag	cagcagcgcc	3540
agccagctga	gcgcggccct	cctgaaggcc	acccgtccac	ccacccacgt	gagccccggc	3600
gcccacgtga	tcgaggccaa	cctgtgttgg	cgccaggaga	tggggcgcaa	catcaccggc	3660
gtgggagagcg	agaacaaggt	ggtggtgcgt	gacagcttcg	acccctgtcg	cgccgaggag	3720
gacgagcgcg	aggtgagcgt	gcccggcgag	atctgtgcga	agagcaagaa	gttccccggc	3780
gccatgcca	tctggggccc	cccccgtact	aaacccccc	tgctggagag	ctggaaaggac	3840
cccgactacg	tgcggggccgt	ggtgcacggc	tgccccctgc	cccccatcaa	gcccccccccc	3900
atcccccccc	ccgcggccaa	gcccacgtg	gtgtgaccgc	agagcagcgt	gagcagcgcc	3960
ctggccgagc	tggccaccaa	gaccttcggc	agcagcgaga	gcagcgcgt	ggacagcgcc	4020
accggccaccg	ccctgccccg	ccaggccaggc	gacgcggcg	acaagggcag	cgacgtggag	4080
agctacagca	gcatggccccc	cctggggg	gagccccggc	accccgacat	gagcgcacggc	4140
agctggagac	ccgtgagcga	ggaggccaggc	gaggacgtgg	tgtgtgcgt	catgagctac	4200
acctggaccg	ggcccccgtat	cacccctgtc	gcccggagg	agagcaagct	gcccatcaac	4260
gccctgagca	acagcctgt	gcccacccac	aaatgggt	acgcacccac	cagccgcagc	4320
gccggcctgc	gccagaagaa	ggtgacctt	gaccgcctgc	aggtgtgttgc	cgaccactac	4380
cgcgacgtgc	tgaaggagat	gaaggccaag	gccagcaccc	tgaaggccaa	gctgtgagc	4440
gtgggaggagg	cctgcaagct	gacccccc	cacagcgcca	agagcaagtt	cggtctacggc	4500
gccaaggacg	tgcgaacact	gaggcagaag	gcccgtgaa	acatccacag	cgtgtggaa	4560
gacctgtgg	aggacaccgt	gacccccc	gacaccacca	tcatggccaa	gaacggaggtg	4620
ttctgcgtgc	agcccgagaa	ggggggccgc	aagccccc	gcctgtatgt	gttccccggc	4680
ctggggcgtgc	gcgtgtgcga	gaagatggc	ctgtacgacg	tgggtggac	cctggggccag	4740
gtgggtatgg	gcagcagcta	cgggtttccag	tacagcccc	gccagcgtgt	ggagttccctg	4800
gtgaacacct	ggaagagcaa	aaagaacccc	atgggtttca	gctacgacac	ccgctgtttc	4860
gacagcaccg	tgaccggaaa	cgacatccgc	gtggaggaga	gcatctacca	gtgtgcgcac	4920
ctggggcccg	aggcccccca	ggccatcaag	agccgtaccc	agccctgtat	catcgccggc	4980
ccccctgacca	acagcaaggg	ccagaactgc	ggcttaccgc	gctggccgc	cagccggcgt	5040
ctgaccacca	gctggggcaa	caccctgacc	tgttacccgt	aggccacgc	cgcctgcccgc	5100
gccgccaaggc	tgcaggactg	caccatgt	gtgaacgcgc	ccggcctgtt	gggtatctgc	5160
gagagcggccg	gcaccccgaaa	ggacggccccc	agccgtccgc	tgttacccga	ggccatgacc	5220
cgctacagcg	ccccccccgg	cgacccccc	cagccccgt	acgacccgtt	gctgtatcacc	5280
agctgcagca	gcaacgtgag	cgtggccccc	gacgcgcgt	gcaagcgcgt	gtactacctg	5340
acccgcgacc	ccaccacccc	cctggccccc	gcccctggg	agaccgcggc	ccacaccccc	5400
gtgaacacgt	ggctgggcaa	catcatcatg	tacccccc	ccctgtgggc	ccgcacatgtc	5460

ctgatgaccc	acttcttcag	catcctgctg	gcccaggagc	agctggagaa	ggccctggac	5520
tgccagatct	acggcgctg	ctacagcatc	gagccctgg	acctgcccc	gatcatcgag	5580
cgccctgcacg	gcctgagcgc	cttcagctg	cacagctaca	gccccggcg	gatcaaccgc	5640
gtggccagct	gcctgcgcaa	gctgggctg	ccccccctgc	gcgtgtggcg	ccaccgcgccc	5700
cgcagcgtgc	gcgcccgcct	gctgagccag	ggccggcccg	ccgcccacctg	cggcaagtac	5760
ctgttcaact	gggcccgtgaa	gaccaagctg	aagctgaccc	ccatccccgc	cgccagccag	5820
ctggacctga	gcggctggtt	cgtggccggc	tacagcggcg	gcgacatcta	ccacagcctg	5880
agccgcgccc	gcccccgctg	gttcatgctg	tgctgagcgt	gggggtgggc		5940
atctacctgc	tgcccaaccg	ctaaa				5965

<210> 4

<211> 37090

<212> DNA

<213> Artificial Sequence

<220>

<223> MRKAd6-NSmut nucleic acid

<400> 4

catcatcaat	aatataccctt	attttggatt	gaagccaata	tgataatgag	gggggtggagt	60
ttgtacgtg	gchgccccgg	tggaacgggg	gcgggtgacg	tagtagtgtg	gcgaaagtgt	120
gatgttgc	gtgtggcgga	acacatgtaa	gcgacggatg	tggaaaagt	gacgtttttg	180
gtgtgcggc	gtgtacacag	gaagtgcaca	tttgcgcg	gttttagcg	gatgtttag	240
taaatttgg	cgttaaccgg	taagatttgg	ccatccgc	ggggaaactg	aataagagga	300
agtgaatct	gaataatttt	gtgttactca	tagcgcgtaa	tatttgcata	ggggcgcggg	360
gactttgacc	gtttacgtgg	agactcgccc	aggtgttttt	ctcaggtgtt	ttccgcgttc	420
cgggtcaaag	ttggcggtttt	attattatag	gcggccgcga	tccattgcatt	acgttgtatc	480
catatcataa	tatgtacatt	tatattggct	catgtccaaac	attaccggca	tgttgacatt	540
gattattgac	tagttattaa	tagaatcaa	ttacggggtc	attagttcat	agccatata	600
tggagttccg	cgttacataa	cttacggtaa	atggccgcgc	tggctgaccc	cccaacgacc	660
cccgccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	720
attgacgtca	atgggtggag	tatattacgg	aaactgcaca	cttggcagta	catcaagtgt	780
atcatatgcc	aagtacgccc	cctattgcgc	tcaatgacgg	taaatggccc	gcctggcatt	840
atgcccagta	catgaccctt	tgggacttgc	ctacttggca	gtacatctac	gtattagtc	900
tcgcttattac	catggtgatg	cgggttttggc	agtacatcaa	tgggcgttga	tagcgggttgc	960
actcacgggg	atttcaatgt	cttccacccca	ttgacgtaa	tgggagtttgc	ttttggcacc	1020
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccatggacg	caaatggcgc	1080
gtaggcgtgt	acgggtggag	gtctatataa	gcagagctcg	tttagtgaac	cgtcagatcg	1140
cctggagacg	ccatccacgc	tgttttgcacc	tccatagaag	acacccggac	cgatccagcc	1200
tccgcggccg	gaaacgggtgc	atttggaaacgc	ggattccccc	tgccaagagt	gagatctgcc	1260
accatggcgc	ccatcacggc	ctactccaa	cagacgcggg	gcctacttgg	ttgcatcatc	1320
actagcctta	caggccggga	caagaaccag	gtcgagggag	aggttcagg	gtttccacc	1380
gcaacacaat	ccttccttgc	gacctgcgtc	aacggcgtgt	gttggaccgt	ttaccatgtt	1440
gctggctcaa	agacctttagc	cggcccaaag	gggcaatca	cccaagatgt	cactaatgt	1500
gaccagggacc	tcgtcggtcg	gcaggcgccc	cccgccggcgc	gttccttgac	accatgcacc	1560
tgtggcagct	cagaccttta	cttggtcacg	agacatgtcg	acgtcatcc	ggtgcggccgg	1620
cggggcgaca	gtagggggag	cctgtctcc	cccaggccgt	tctccatctt	gaagggctct	1680
tcgggtggtc	cactgtctcg	cccttcgggg	cacgtgtgg	gcatcttccg	ggctgcccgt	1740
tgcacccggg	gggttgcgaa	ggcgggtggac	tttgcggcc	taggtccat	ggaaactact	1800
atgcgggttc	cgttcttcac	ggacaactca	tcccccccg	ccgttaccgca	gtcatattcaa	1860
gtggccacc	tacacgctcc	cactggcagc	ggcaagagta	ctaaagtgcc	ggctgcata	1920
gcagcccaag	ggtacaagg	gttcgttcc	aatccgtcc	ttggccgtac	ctttagggttt	1980
ggggcgata	tgtctaaggc	acacggatt	gaccccaaca	tcaagaactgg	ggtaaggacc	2040
attaccacag	gcccggccgt	cacatactct	acctatggca	agtttcttgc	cgatgggtgt	2100
tgctctgggg	gcgcttatga	catcataata	tgtatgagt	gccattcaac	tgactcgact	2160

acaatcttgg	gcatcggcac	agtccctggac	caagcggaga	cggctggagc	gcggcttgc	2220
gtgctcgcca	ccgctacgccc	tccgggatcg	gtcaccgtgc	cacacccaaa	catcgaggag	2280
gtggccctgt	ctaatactgg	agagatcccc	ttctatggca	aagccatccc	cattgaagcc	2340
atcagggggg	gaaggcatct	cattttctgt	cattccaaga	agaagtgcga	cgagctcgcc	2400
gcaaagctgt	caggcctcg	aatcaacgct	gtggcgatt	accggggct	cgatgtgtcc	2460
gtcataccaa	ctatcgaga	cgtcgttgc	gtggcaacag	acgctctgt	gacgggctat	2520
acggggcact	ttgactcgat	gatcgactgt	aacacatgtg	tcacccagac	agtcgacttc	2580
agcttggatc	ccacacccac	cattgagacg	acgaccgtgc	ctcaagacgc	agtgtcgccg	2640
tcgcagcggc	ggggtaggac	tggcaggggt	aggagaggca	tctacagggt	tgtactccg	2700
ggagaacggc	cotcgggcat	tttcgattcc	tcggctctgt	gtgagtgcta	tgacgcgggc	2760
tgtgcttgggt	acagactcac	ccccggccag	acctcggtt	gggtcgccgc	ctacccatgg	2820
acaccagggt	tggccgtttg	ccaggaccac	ctggagttt	gggagagtgt	cttcacaggc	2880
ctcaccacaca	tagatgcaca	tttcttgc	cagaccaagc	aggcaggaga	caacttcccc	2940
taccttggtag	cataccaaagc	cacgggtgc	gccaggcgc	aggccccacc	tccatcatgg	3000
gatcaaatgt	ggaagtgtct	catacgctg	aaacctacgc	tgacacgggc	aacacccttgc	3060
ctgtacaggc	tgggagccgt	ccaaaatggag	gtcaccctca	cccacccat	aaccaaaatac	3120
atcatggcat	gcatgtcgcc	tgaccttggag	gtcgtacta	gcacccgggt	gctgggtggc	3180
ggagtccttg	cagctctggc	cgcgtattgc	ctgacaacag	gcagtgtgt	cattgtgggt	3240
aggattatct	tgtccggag	gcccggctatt	gttcccgaca	gggagtttct	ctaccaggag	3300
ttcgatgaaa	tggaagagtg	cgccctcgac	ctcccttaca	tcgagcagg	aatgcagctc	3360
gcccggcaat	tcaagcggaa	agcgctcggg	ttactgcaaa	cagccaccaa	acaagcggag	3420
gctgtcgctc	ccgttgggaa	gtccaaatgg	cgagcccttg	agacattctg	ggcgaagcac	3480
atgttggaaatt	tcatcagcgg	gatacagttac	tttagcggct	tatccactt	gcctgggaac	3540
cccgcaatag	catcattgtat	ggcattcaca	gccttatca	ccagcccgct	caccacccaa	3600
agtaccctcc	tgttaacat	cttggggggg	tgggtggctg	cccaactcgc	cccccccgac	3660
gccgttccgg	ctttctgtggg	cgccggcata	gccgggtgcgg	ctgttggcag	cataggcctt	3720
gggaagggtgc	ttgtggacat	tctgggggt	tatggagcag	gagtggccgg	cgcgctcg	3780
gccttcaagg	tcatgagcgg	cgagatccc	tccaccgagg	acctggtaa	tctacttcc	3840
gccatccctt	ctccttggcgc	cctggctg	gggtcgctgt	gtcagcaat	actgcgtcg	3900
cacgtgggtc	cgggagaggg	ggctgtgcag	tggatgaacc	ggctgataac	gttcgcctcg	3960
cggggtaatc	atgtttcccc	cacgactat	gtgcctgaga	gacgcggcc	agcgcgtgtt	4020
actcagatcc	tctccagect	tacatcact	cagctgtga	aaaggctcca	ccagttggatt	4080
aatgaagact	gtccacacc	gttccggc	tcgtggctaa	gggatgtttt	ggactggata	4140
tgcacgggt	tgactgactt	caagactgg	ctccagttca	agctctcg	gcagctacc	4200
ggagtcctt	tttctctgt	ccaacgggg	tacaaggag	tctgggggg	agacggcatc	4260
atgcaaacca	cctgcccatt	tggagcacag	atcacccggac	atgtcaaaa	cggttccatg	4320
aggatcgctg	ggcctaagac	ctgcagcaac	acgtggcatg	gaacattccc	catcaacgca	4380
tacaccacgg	gccccctgcac	accctctcca	gcccacaaact	attctagggc	gctgtggcgg	4440
gtggccgtg	aggagtacgt	ggaggtcacg	cgggtgggg	atttccacta	cgtgacggc	4500
atgaccactg	acaacgtaaa	gtcccatgc	caggttccgg	cttcctgaatt	cttcacggag	4560
gtggacggag	tgcggttgc	caggtacgct	ccggcgtgc	ggcctctct	acgggaggag	4620
gttacattcc	aggtcgccgt	caaccaatac	ctgttgggt	cacacttacc	atgcgagcccc	4680
gaacccggatg	tagcagtgt	cacttcatg	ctcaccggacc	cctccacat	cacacggacaa	4740
acggctaaagc	gttaggttggc	cagggggtct	ccccccctt	tgccagctc	ttcagctagc	4800
cagtgtctg	cgcccttctt	gaaggcgaca	tgcaactaccc	accatgttc	tccggacgct	4860
gacccatcg	aggccaaacct	cctgtggcg	caggagatgg	gcccggacat	caccccgctg	4920
gagtcggaga	acaagggtgtt	agtccctggac	tcttcgacc	cgcttcg	ggaggaggat	4980
gagagggaaag	tatccgttcc	ggcggagatc	ctgcggaaat	ccaagaat	cccccccg	5040
atgcccattt	ggggcgcccc	ggattacaac	cctccactgt	tagactctg	gaaggacccg	5100
gactacgtcc	ctccgggtgt	gcacgggtgc	ccgttgcac	ctatcaaggc	cccttccaaata	5160
ccacccatccac	ggagaaagag	gacgggtgtc	ctaacagatgt	cctccgtgtc	ttctgcctt	5220
gcggagctcg	ctactaagac	cttcggcagc	tccgaatcat	cgccgtcg	cagccgcacg	5280
gcgaccggccc	ttccctgacca	ggcctccgac	gacgggtgaca	aaggatccga	cggtgagtc	5340
tacttccca	tgccccccct	tgagggggaa	ccgggggacc	ccgatctcag	tgacgggtct	5400
tggtctaccg	tgagcgagga	agctagtgt	gatgtcg	gctgctcaat	gtccatcaca	5460

tggacaggcg	ccttgatcac	gccatgcgct	gcggaggaaaa	gcaagctgcc	catcaacgcg	5520
tttagcaact	ctttgctcg	ccaccataac	atggttatg	ccacaacatc	tcgcagcgca	5580
ggcctgcggc	agaagaaggt	caccttgcac	agactgcaag	tcctggacga	ccactaccgg	5640
gacgtgctca	aggagatgaa	ggcgaaggcg	tccacagtt	aggctaaact	cctatccgt	5700
gaggaagcct	gcaagctgac	gccccccat	tcggccaaat	ccaagttgg	ctatggggca	5760
aaggacgtcc	ggaacctatc	cagcaaggcc	gttaaccaca	tccactccgt	gtggaggac	5820
ttgctggaag	acactgtgac	accaattgac	accaccatca	tggcaaaaaa	tgaggtttc	5880
tgtgtccaac	cagagaaagg	aggccgtaa	ccagccgc	ttatcgatt	cccagatctg	5940
ggagtccgt	tatgcgagaa	gatggccctc	tatgatgtgg	tctccaccc	tcctcaggtc	6000
gtgatggct	cctcatacg	atccagtagc	tctctgggc	agcgagtcg	gttctggtg	6060
aatacctgga	aatcaaagaa	aaaccccatg	ggcttttcat	atgacatcg	ctgtttcgac	6120
tcaacggtca	ccgagaacga	catccgttt	gaggagtc	tttaccaatg	tttgacttg	6180
gccccccaa	ccagacaggc	cataaaatcg	cteacagac	ggctttatat	cgggggtcct	6240
ctgactaatt	caaaaaggcg	gaactgcccgt	tatccgggt	gccgcgcgag	cggcgtgctg	6300
acgactagct	gcccgttac	cctcacatgt	tacttgaagg	cctctgcagc	ctgtcgagct	6360
gcaagctcc	aggactgac	gatgctcgt	aacgcgcgg	gccttgcgt	tatctgtaa	6420
agcgcggaa	cccaagagga	cgcgccgagc	ctacgagtct	tcacggaggc	tatgactagg	6480
tactctggcc	ccccccggga	cccgcccaa	ccagaatacg	acttggagct	gataacatca	6540
tgttcccca	atgtgtcggt	cgccccacgt	gcatcaggca	aaaggggtgt	ctacactcacc	6600
cgtgatccca	ccacccccc	cgcacgggt	gcgtggggaa	cagctagaca	cactccagtt	6660
aactcctggc	taggcaacat	tatcatgtat	gcgcccactt	tgtggcaag	gatgattctg	6720
atgacttca	tcttctccat	ccttctagca	cagagcaac	ttgaaaaaagc	ccttggactgc	6780
cagatctacg	gggcctgtta	ctccatttag	ccacttgacc	tacccatcgat	cattgaacga	6840
ctccatggcc	ttagcgcatt	ttcaacttcat	agtacttctc	caggtgagat	caatagggtg	6900
gcttcatgca	tcagggaaact	tgggttacca	cccttgcgag	tctggagaca	tcggccagg	6960
agcgtccg	ctaggtact	gttccagggg	gggaggggcg	ccacttgcgg	caagttacctc	7020
ttcaacttgg	cagtgaagac	caaactaaa	ctcaactccaa	tcccgctgc	gtcccagtc	7080
gacttgcgg	gctgggtcgt	tgctgggtac	agccccggag	acatataatca	caggctgtct	7140
cgtccccac	cccgctgggt	catgctgtc	ctactctac	tttctgttag	ggtaggcata	7200
tacctgtcc	ccaaaccggta	aatctagagc	tgtgccttct	agttgccagc	cattctgtgt	7260
ttggccctcc	cccgctgcct	ccttgaccct	ggaaggtgcc	actcccaactg	tcctttccct	7320
ataaaaatgag	gaaattgtat	cgcattgtct	gagtaggtgt	cattctattc	tgggggttgg	7380
ggtggggcag	gacagcaagg	gggaggattg	ggaagacaat	agcaggcatg	ctggggatgc	7440
ggtgggtct	atggccgatc	ggcgcggcgt	actgaaatgt	gtgggcgtgg	cttaagggtg	7500
ggaaagaata	tataagggtgg	gggtctttag	tagtttgt	tctgttttgc	agcagccgc	7560
gccgcctatga	gcacccaactc	gtttgtatgg	agcattgtg	gtctatattt	gacaacgcgc	7620
atggccctat	ggggccggggt	gctgtcagaat	gtgtgggtct	ccagcattga	tggcgtcccc	7680
gtcctggcc	caaaactctac	taccttgacc	tacgagaccg	tgtctggaaac	gcccttggag	7740
actgeagcc	ccggccggc	ttcagccgt	gcagccaccc	cccgccggat	tgtactgac	7800
tttgccttcc	tgagcccgct	tgcaagcagt	gcagcttccc	gttcatccgc	ccgcgatgac	7860
aagttgacgg	cttttttggc	acaatttggat	tcttgaccc	ggaaacttaa	tgtcgtttct	7920
cagcagctgt	tgtatctgc	ccagcagggt	tctgccttgc	aggcttccctc	ccctcccaat	7980
gccccctaaa	acataaataaa	aaaaccagac	tctgttttgg	tttggatcaa	gcaagtgtct	8040
tgctgtcttt	atttaggggt	tttgcgcg	cggtaggccc	gggaccagcg	gtctcggtcg	8100
tttgggggtcc	tgtgtatttt	ttccaggacg	tgttaaaggt	gactctggat	gttcagatac	8160
atgggcataa	gcccgtctc	gggggtggagg	tagcaccact	gcagagcttc	atgtcgccgg	8220
gtgggtttgt	agatgtatca	gtcgttagc	gagcgttggg	cgtgtgcct	aaaaatgtct	8280
ttcagtagca	agctgattgc	cagggccagg	cccttgggt	aagtgtttac	aaagcggta	8340
agctgggtat	ggtgcatacg	tggggatatg	agatgtatct	tggactgtat	ttttaggttg	8400
gctatgtcc	cagccatatac	cctccgggg	ttcatgttgc	gcagaaccac	cagcacatgt	8460
tatccgggtc	acttggggaa	tttgcatgt	agctttagaa	gaaatgcgt	gaagaacttg	8520
gagacggcc	tgtgaccc	aagatttcc	atgcattcg	ccataatgt	gcaatgggc	8580
ccacggccgg	cggcctgggc	gaagatattt	ctggatcac	taacgtcata	gttgcgttcc	8640
aggatgagat	cgtcatagc	cattttaca	aagcgcggc	ggagggtgcc	agactgcgt	8700
ataatggttc	catccggccc	agggcgtag	ttaccctac	agatttgcat	ttccacgcgt	8760

ttgagttcag atgggggat catgtctacc tgcggggcga tgaagaaaac ggtttccggg	8820
gtagggaga tcagctgggaa agaaaggcagg ttcctgagca gctgcgactt accgcagccg	8880
gtggggccgt aaatcacacc tattaccggc tgcaactggg agttaagaga gctgcagctg	8940
ccgtcatccc tgagcagggg ggcacttcg ttaagcatgt ccctgactcg catgtttcc	9000
ctgaccataat ccgcccagaag ggcactccggc cccagcgata gcagttctg caaggaagca	9060
aagttttca acgggtttagg accgtccggc gtaggcattgc tttttagcg ttagccaagc	9120
agttccaggg ggtcccacag ctgggtcacc tgctctacgg catctcgatc cagcatatct	9180
cctcggttcg cgggttgggg cgggttcg ttagacggcag tagtcggtgc tcgtccagac	9240
gggcagggt catgtcttc cacggggcga gggtcctcg cagcgtagtc tgggtcacgg	9300
tgaagggtg cgctccggc tgcgcgtgg ccagggtgcg cttgaggctg gtccgtctgg	9360
tgctgaagcg ctgcgggtct tgcctcgcc cgtagggccag gttagcatttgc accatgggt	9420
catagtcagg cccctccggc gctgtggccct tggcgcgcag cttggccctt gaggaggcgc	9480
cgcacgaggc gcaactggcaga ctttgaggg ctagagatgg gggcgcgaga aataccatt	9540
ccggggagta ggcattccggc cgcaggcccc cgcagacggt ctcgcatttc acgagccagg	9600
ttagctctgg cgctccggg tcaaaaacca gtttcccccc atgcttttgc atgcgtttct	9660
tacctctgtt ttccatgagc cgggttccac gctcggtgac gaaaaggctg tccgtgtccc	9720
cgtatacaga cttgagaggg ctgtctcgat gctgtgttcc gctgttcc tctgtatagaa	9780
actcggacca ctctgagacg aaggctcgcc tccaggccag cacgaaggag gctaagtggg	9840
aggggttagcg gtcgttgcact aactaggggtt ccactcgctc cagggtgtga agacacatgt	9900
cgccttcggc ggcattcaagg aaggtgattt gttttaggtt gtaggcacg tgaccgggtg	9960
ttcctgaagg ggggtataa aagggggtgg gggcgcgttc gtcctcaactc tcttccgcatt	10020
cgctgtctgc gaggggccagg tgggggttgc agtactccct ctcaaaaacgc ggcattactt	10080
ctgcgctaag attgtcagg tccaaaacacg aggaggattt gatatttcc tggcccgcgg	10140
tgtatccctt ggggtggcc gcttccatctt ggtcagaaaaa gacaatctt ttgttgcataa	10200
gctgtgtggc aaacgaccgg tagagggcgt tggacagcaa cttggcgatg gagcgcagg	10260
tttgggggg tgcgcgttc ggcgcgttcc tggccgcgtat gtttagctgc acgtattcgc	10320
gchgcaacgca cgcatttcg gggaaagacgg tggcgctc gtcgggcaact aggtgcacgc	10380
gcacaccgcg gttgtgcagg gtgacaagg tcaacgctgtt ggctacactt ccgcgttaggc	10440
gctcggttgtt ccagcagagg cggccccc tgcgcgagca gaatggcggt agtgggtcta	10500
gctgcgtctc gtccgggggg tctgcgttca cggtaaagac cccgggcacg aggccgcgt	10560
cgaagtagtc tatcttgcatt ctttgcaggat ctaggcctcg ctgcgcattcg cggggcgcaaa	10620
gchgccgcgtc gtatgggttgc agtgggggat cccatggcat ggggtgggttgc acgcggagg	10680
cgtacatgcc gcaaatgtcg taaacgtaga ggggttctctt gatatttcca agatatgttag	10740
ggtagcatctt tccaccgcgg atgtgcgc gcaactatc gtagatgtcg tgcgaggggag	10800
cgaggaggc gggaccggagg ttgctacggg cggcgctc tgcgcggaaactatctgc	10860
tgaagatggc atgtgagttt gatgatatgg ttggacgctg gaagacgtt aagctggcg	10920
ctgtgagacc taccgcgtca cgcacgaagg aggctgttgg gtcgcgcacg ttgttgcacca	10980
gctcgccgtt gacctgcacg tctagggcgc agtagtccag gttttcttgc atgatgtcat	11040
acttatectt tccctttttt ttccacagct cgcgttgcgg gacaactt tgcggctt	11100
tccagttactc ttggatcgga aaccctcggtt cttccgcacg gtaagggactt agcatgtt	11160
actgggttgcac ggccttggtag ggcgcacatc cttttctac gggtagcgccg tatgcctgcg	11220
cgcccttccg gaggcgagggtt tgggtgagcg caaagggttgc ctttccatcg actttggat	11280
actggattt gaaatgtcgatc tgcgcgttcc cggccctgc tccatcgatc aacttcccg	11340
gcttttggaa acgggggtt ggcaggcgca aggtgacatc gtttagatgc actttcccg	11400
cgcgaggcat aaatgtcgatc tgatgcggg aggggtcccg cacctcgaa cgggtgtt	11460
ttacccggc ggcgaggatc atctcgatc acccggttgc gttgtggcc acaatgtt	11520
gttccaagaa ggcggggatc ccctgtatgg aaggcaattt tttaaatgtcc tgcgttgc	11580
gcttttgcagg ggagctgacg cgggtgttcc aagggccca gtcgttgc gtaggggttgg	11640
aaggcgttac gtagtccac aggttgcacggg ccattagcat ttgcagggttgc gtcgaaagg	11700
tcctaaactt ggcacatcgatc gccattttt ctgggtgtat gcaatgtt gtaagcggtt	11760
cttgcgttccca ggggtccat ccaaggcccg cggcttaggtc tgcgcgcgc gtcactatcg	11820
gctcatctcc ggcgttccatc atgaccatcgatc tgaaggccac gagctgttcc cccaaaggccc	11880
ccatccaatgtt ataggtctctt acatcgatgg tgcacaaaggac acgttgcgttgc cgaggatgcg	11940
agccatcgatgg gaaaggactgg atctcccgcc accagttgg gtaggtggctt tgatgttgg	12000
gaaatgtt gtcctgcgat cggccgttac actcgatgttgc gttttgtt gaaacgtgcgc	12060

agtactggca	gcgggtgcacg	ggctgtacat	cctgcacgag	gttgacctga	cgaccgcgca	12120
caaggaagca	gagtggaaat	ttgagccct	cgccctggcg	gtttggctgg	tggtcttcta	12180
cttcggctgc	ttgtccttga	ccgtctggct	gctcgagggg	agttacggtg	gatcgacca	12240
ccacgcccgg	cgagcccaa	gtccagatgt	ccgcgcgcgg	cggtcgaggc	ttgtatgacaa	12300
catcgccgag	atgggagctg	tccatggct	ggagctcccg	cggtcgctgg	tcaggcggga	12360
gctccctgcag	gtttacctcg	catagccgg	tcagggcgcg	ggcttaggtc	agggtatacc	12420
tgatattccg	gggctgggtt	gtggcggcgt	cgatggctt	caagaggccg	catcccccg	12480
gcccggactac	ggtacccgccc	ggcggggcggt	ggggccgggg	ggtgtccctt	gatgtatgc	12540
ctaaaagcgg	tgacccgggc	ggggcccccgg	aggtaggggg	ggctcgggac	ccggccgggg	12600
agggggcagg	ggcacgtcg	cgccgcgcgc	ggcaggagac	ttgtgtcg	cgccggagggtt	12660
gctggcgaac	gcaacgcacgc	ggcggttgc	ctccctgaaatc	ttggcgcctt	gcgtgaagac	12720
gacggggccg	gtgagcttga	acctgaaaga	gagttcgaca	aatcaattt	cggtgtcg	12780
gacggggcc	ttggcgaaaa	tccctgcac	gtccctgtag	ttgtcttgc	aggcgatctc	12840
ggccatgaac	tgctcgatct	cttccctctg	gagatctccg	cgtccggc	gtccacgggt	12900
ggcggcgagg	tcgttggaga	tgcggccat	gagctgcgag	aaggcg	ggccctccctc	12960
gttccagacg	cggtctgtaga	ccacgcccc	tccgcacatcg	cgggcgcgc	tgaccacctg	13020
cgcgagattg	agctccacgt	ggcgggcgaa	gacggcgtag	tttcgcaggc	gtctaaagag	13080
gtatgttgg	gtggtggcgg	tgtgttctgc	caccaagaag	tacataaccc	agcgcgcgaa	13140
cgtggattcg	ttgtatatccc	ccaaggcctc	aaggcgctcc	atggcctcg	agaatgtccac	13200
ggcgaagttg	aaaactggg	agttgcgc	cgacacgggt	aactccttct	ccagaagacg	13260
gatgagctcg	gcaacgtgt	cgccgcac	cgctcaaaag	gctacagggg	cctttcttc	13320
tttccaaatc	tccttttcca	taagggcctc	cccttcttct	tcttcgcg	cggttggggg	13380
aggggggaca	cgccggcgac	gacggcgcac	cgggaggcg	tcgaca	aaagcgtcat	13440
ctcccccgg	cgacggcgca	tggtctcg	gacggcgcgg	ccgttctcg	ggggggcg	13500
ttggaagacg	ccgccccgtca	tgtcccggtt	atgggttggc	ggggggctgc	cgtcggcg	13560
ggatacggcg	ctaacgatgc	atctcaacaa	tttttgtgt	gttactccgc	caccgaggga	13620
cctgagcgag	tccgcacatcg	ccggatcg	aaacctctcg	agaaaggcg	ctaaccagtc	13680
acagtccaa	ggtaggctga	gcaccgtgg	ggcggcgcag	ggcggcggt	cggggttgg	13740
tctggccgag	gtgctgtga	tgtgttatt	aaagttaggc	gtcttgcg	ggccgatgg	13800
cgacagaagc	accatgtcct	ttgggtccggc	ctgtgtatg	cgcaggcggt	cgccatgc	13860
ccaggcttcg	ttttgcacatc	ggcgcagggt	ttttgtatg	tcttcgc	gccttttctac	13920
cggcacttct	tcttcgcctt	ccttcgttgc	tgcacatctt	gcacatctcg	ctgcggcg	13980
ggcggagttt	ggccgttaggt	ggccgcctt	tcttcgcat	cgtgtgaccc	cgacgcctt	14040
catcggtga	agcaggccca	ggtcggcgac	aaacgcgtcg	gctaataatgg	cctgtgcac	14100
ctgcgtgagg	gtagacttga	agtcgttccat	gtccacaaaag	cggtggat	cgcccggtgt	14160
gatgggtta	gtgcagttgg	ccataacgg	ccagtttacg	gtctggat	ccggctgcga	14220
gagctcggt	tacctgagac	cgagtaagc	ccttgcgtca	aagacgtat	cgttgc	14280
ccgcaccagg	tactggatc	ccaccaaaaa	gtgcggcg	ggctggcggt	agagggggcca	14340
gcttagggtg	gccggggctc	cgggggcgag	gttttcaac	ataaggcgat	gatatccgt	14400
gatgtacccgt	gacatccagg	tgtatgcggc	ggcgggtgg	gaggcgcgcg	gaaagtca	14460
gacgcgttcc	cagatgttgc	gcagcgc	aaagtgc	atggtgcgg	cgctctgg	14520
ggtcaggcg	gcaacgtcg	tgacgc	gaccgtgc	aaggagagc	tgtaaagcgg	14580
caactctccg	ttgtctgg	gataaaattcg	caagggtatc	atggcgg	acccgggtt	14640
gaaccccg	tcggccgtc	cgccgtatc	catgcgg	ccggccgcgt	gtcgaaccca	14700
ggtgtcg	gtcagacaac	ggggggagc	tccttttgc	ttcccttc	cgccggcg	14760
tgctcgct	gttttttgc	ccactggcg	cgccgcgcgt	aagcggtt	gtctggaa	14820
gaaagcatta	agtggctcg	tccctgtac	cgagggtt	ttttcaagg	tttgagatcg	14880
gggacccccc	gttcgagtct	cgggccggc	ggactgcgg	gaaacggggt	ttgcctcccc	14940
gtcatgc	agccccgtt	caaatttctc	cgaaaacagg	gacgagcccc	ttttttgtct	15000
ttcccaatgt	catccgggtc	tgccgc	aggccccct	cctcagc	ggcaagagca	15060
agagcagcgg	cagacatgc	gggcaccc	cccttcttct	accgcgtc	gaggggca	15120
atccgcggct	gacgcggcgg	cagatgt	ttacgaaccc	ccgcggc	ggacccgg	15180
ctacttggac	ttggaggagg	gcaaggccct	ggcgcggct	ggagcgc	ctccgtgc	15240
acaccaagg	gtcagctga	agcgtgac	gca	cgaggcg	ggcagaac	15300
gttcgcgac	cgcgaggagg	aggagcc	ggagatgc	gatcgaa	gttccatgc	15360

gcgcgagttg cggcatggcc tgaaccgcga gcgggttgcg	cgcgaggagg actttgagcc	15420
cgacgcgccc accgggatta gtcccgccg cgacacacgtg	gcggccccc acctggtaac	15480
cgcgtagcag cagacggta accaggagat taactttcaa	aaaagcttta acaaccacgt	15540
gcccacgctt gtggcgccg aggagggtgc tataggactg	atgcacatgt gggactttgt	15600
aagcgcgtg gagcaaaaacc caaatagcaa gcegcctatg	gcegcgtgt tccttatagt	15660
gcagcacagc agggacaacg aggcatcag ggatgcgtg	ctaaacatag tagagcccga	15720
ggccgctgg ctgctcgatt tgataaacat tctgcagagc	atagtggtc aggagcgcag	15780
cttggacctg gctgacaagg tggccgcccata taactattcc	atgctcagtc tggcaagtt	15840
ttaccccgc aagatataacc ataccctta cgttcccata	gacaaggagg taaagatcga	15900
ggggttctac atgcgcatgg cgctgaaggt gcttaccttgc	agcgacgacc tggcgcttta	15960
tcgcaacgag cgcacatccaca aggccgtgag cgtgacccgg	cgccgcgacgc tcagcgaccg	16020
cgagctgtatc cacagcctgc aaaggcccctt ggctggcacg	ggcagcgcgcg atagagaggc	16080
cgagcttac ttgacgcgg ggcgtgacccgttggcc	ccaagccgcac ggcgccttgg	16140
ggcagctggg gccggacccgtt ggctggcggtt ggaccccg	cgcgctggca acgtcgccgg	16200
cgtggaggaa tatgacgagg acgtgagta cgagccagag	gacggcgagt actaagcggt	16260
gatgtttctg atcagatgtatc gcaagacgcac accggacccgg	cggtgcgggc ggcgcgtcag	16320
agccagccgtt cccgccttaa ctccacggac gactggcgcc	aggtcatggaa ccgcacatcatg	16380
tcgctgactg cgcgcaacccc tgacgcgttc cgccagcagc	cgcaggccaa ccggctctcc	16440
gcaattctgg aagcggtggt cccggcgcgc gcaaacccttca	cgcacagaaa ggtgtggcg	16500
atcgtaaacg cgctggccga aaacaggggcc atccggcccg	atgaggccgg cctgttctac	16560
gacgcgctgc ttcaagcgatc ggctcgatc aacagcagca	acgtgcagac caactggac	16620
cggtcggtgg gggatgtgcg cgaggccgtt ggcgcgtgt	agcgcgcgc gCACAGGGC	16680
aacatgggctt ccatgggttgc actaaacgc ttctgtata	caacgtggccg	16740
cggggacagg aggactacac caactttgtt acgcgcactgc	ggctaattgt gactgagaca	16800
ccgcggaaatgtt aggtgtatca gtccggccca gactatttt	tccagaccag tagacaaggc	16860
ctgcagaccg taaaccttagt ccaggcttcc aagaacttgc	aggggctgtt ggggggtgg	16920
gctcccacag ggcacccgcg gaccgtgtct agcttgcgt	cgcccaactc ggcgcgttgc	16980
ctgctgctaa tagcgccctt cacggacagt ggcagcgtgt	ccggggacac ataccttaggt	17040
cactgtgtca cactgttaccg cgaggccata ggtcaggcgc	atgtggacga gcataactttc	17100
caggagatta caagtgttagt ccgcgcgtgc gggcaggagg	acacggccag cctggaggca	17160
accctgaactt acctgtgtac caacccgcgg caaaaaatcc	cctcggttca cagtttaaac	17220
agcggaggagg agcgcattttt ggcgtatgtt cagcagagcg	tgagcccttgc cctgtatgcgc	17280
gacggggtaa cgcccgacgtt ggcgcgttgcg atgacccgc	gcaacatggaa accgggcgtt	17340
tatgccttca accggccgtt tatcaatgc ctaatggact	acttgcacgc cgccggccgc	17400
gtgaaccccg agtatttac caatgccatc ttgaacccgc	actggcttacc gcccccttgc	17460
ttctacaccg ggggattcga ggtgcccgg ggttaacgtatc	gattcctctg ggacgcacata	17520
gacgacagcg ttttttcccc gcaacccgcg accctgtatc	agttgcaaca acgcgcacag	17580
gcagaggccgg cgctgcgaaa ggaaagttt cgcaggccaa	gcagcttgc cgatctaggc	17640
gctgcggccc cgccgtcaga tgcttagtgc ccatttccaa	gcttgatagg gtcttttacc	17700
agcactcgca ccacccgcg ggcgcgtgt ggcgaggagg	agtacctaa caactcgctg	17760
ctgcagccgc agcgcgaaaaa gaaacctgcctt ccggcgctt	ccaaacaaacgg gataagagac	17820
ctatgtggaca agatgatgtatggaaagacg tatgcgcagg	agcacagggg tggcccccggc	17880
ccgcggccgc ccacccgtcg tcaaaaggc gaccgtcagc	ggggctgtt gttggaggac	17940
gatgacttgcg cagacgacac cagcgcttgc gatggggag	ggagtggca cccgtttca	18000
cacccctgcgc ccaggctgg gagaatgttt taaaaaaaaag	catgtgcac aaaaaaaaaac	18060
tcaccaaggc catggcaccg agcgttgcgtt ttcttgcatt	ccctttagt tgccgcgcgc	18120
ggcgatgtat gaggaaaggc ctcctccctc ctacgagacg	gtgggtgagcg cggcccaagt	18180
ggccggccgc ctgggttac cccttcgtgc tcccctggac	ccggccgtcg tgccctccgc	18240
gtacccgtcg cctaccgggg ggagaaacag catccgttac	tctgagggtt cacccttatt	18300
cgacaccacc cgtgtgtacc ttgtggacaa caagtcaacg	gatgtggcat ccctgaacta	18360
ccagaacgcac cacagcaact ttcttaaccac ggtcattcaa	aacaatgtact acagcccccgg	18420
ggaggcaacg acacagacca tcaatcttgc cgaccggcgtc	cactggggcg ggcacactgaa	18480
aaccatcccg cataccaaca tgccaaatgtt gaaacgttcc	atgtttacca ataagttaa	18540
ggccgggggtt atgggtgtcgc gctcgatc taaggacaaa	caggtggagc tgaaatacga	18600
gtgggtggag ttacgcgtgc ccgaggccaa ctactccgag	accatgacca tagacccat	18660

gaacaacgcg atcgtggagc actacttga a agtgggcagg cagaacgggg ttctggaaag	18720
cgacatcggt gtaaagtgg acacccgcaa cttcagactg gggtttgcacc cagtcactgg	18780
tcttgcata cctggggat atacaaacga agccttccat ccagacatca ttttgcgc	18840
aggatgcggg gtggacttca cccacagccg cctgagcaac ttgttggca tccgcaagcg	18900
gcaacccttc caggaggctt ttaggatcac ctacgatgac ctggagggtg gtaacattcc	18960
cgcactgtt gatgtggacg cctaccaggc aagcttggaa gatgacacccg aacaggcg	19020
gggtggcgca ggccggcgca acaacagtgg cagcggcgcc gaagagaact ccaacgcgc	19080
agctgcggca atgcagccgg tggaggacat gaacgatcat gccattcgcc gcgacaccc	19140
tgccacacgg gcccggaggaga agcgcgtga ggccgaggca gggccgaaag ctggcgcccc	19200
cgctgcggag gctgcacaac ccgaggtcga gaagcctcag aagaaacccg tgattaaacc	19260
cctgacagag gacagcaaga aacgcgtt caacctaata agcaatgaca gcacccctac	19320
ccagtagccg agctggtacc ttgcatacaa ctacggcgac cctcaggccg ggatccgctc	19380
atggaccctg ctttgcactc ctgacgttac ctgcggctcg gaggcgttacttgcgtt	19440
gcccgcacat atgcaagacc ccgtgacccctt cegctccacg cgccagatca gcaacttcc	19500
ggtgggtggc gccgagctgt tgccctgtca ctccaagagc ttctacaacg accaggccgt	19560
ctactcccg ctcatccgccc agtttaccc tctgacccac gtgttcaatc gcttcccg	19620
gaaccagatt ttggcgcccg cgccagcccc caccatcacc accgtcgtt aaaaacgttcc	19680
tgctctcaca gatcacggga cgctaccgtt ggcacacagc atcggaggag tccagcgagt	19740
gaccattact gacgcacagc gccgcacctt cccctacgtt tacaaggccc tggcatagt	19800
ctcgccgcgc gtccatcga gccgcacttt ttgagcaagc atgtccatcc ttatatcgcc	19860
cagcaataac acaggctggg gcctgcgtt cccaaagcaag atgtttggcg gggcaagaa	19920
gctgcgttccgac caacacccag tgcgcgttccg cggcactac cgccgcgcctt gggcgccgca	19980
caaacgcggc cgcaactggc gcaccaccgtt cgatgacgcg atcgcacgcgg tggtggaggaa	20040
ggcgcgcac tacacgcggc cggccgcgc agtgtccacc tgggacgcgg ccattcagac	20100
cgtggcgccg ggagccggc gctacgtttaa aatgaagaga cggcggaggc gcttagcacc	20160
tcggccacccg cggccggccg gcaacttgcgcg cggcggggcc tgcttaaccg	20220
cgcacgtcgc accggccgac gggccggcat cgcacgcgtt cgaaggctgg ccgcgggtat	20280
tgtcaactgtt ccccccaggc ccaggcgacg agccggccgc gcagcagccg cggccattag	20340
tgctatgact cagggtcgca gggcaacgt gtactgggtg cgcactcgg ttagccgcct	20400
gcccgtggcc gtcgcaccc gcccccccgca caactagatt gcaataaaaaa actactttaga	20460
ctcgactgt tttatgttac cagcggccgc ggcgcgcattc gaagctatgt ccaagcgca	20520
aatcaaaagaa gagatgtcc aggtcatcgc gcccggagatc tatggccccc cgaagaagga	20580
agagcaggat tacaagccca gaaagctaaa gcccgttcaaa aagaaaaaga aagatgtga	20640
tgtatgttac cttgacgtac aggtggaaact gttgcacgcg accgcgcggca ggcacccgg	20700
acagtggaaa ggtgcacgcg taagacgtt tttgcgaccc ggcaccaccg tagtctttac	20760
gcccgttgcg cgttccaccc gcacccatca ggcgtgtat gatgagggtgt acggcgacga	20820
ggacctgtt ggcacggccca acgacggccctt cggggaggtt gcttacggaa agccgcataa	20880
ggacatgtctg gcttgcgcg tggacgaggc caacccaaactt cctagcttac agcccggttac	20940
actgcacgcg gtgtggcccg cgttgcacc gtcgttgcggaa aacgcggccca taaagcgccg	21000
gtctgggtac ttggcacccca cctgtcgtt gatgttaccgc aacgcgttac gacttggaa	21060
tgtctttggaa aaaatgaccgc tggagcttgg gcttggggcc gagggtccgc tgccggccat	21120
caagcagggtt gcacccggac tggcgttgc gaccgttggac gttcgttgcatacc caccaccag	21180
tagcaactgtt attgcacttgc ccacagaggc catggagaca caaacgttcc cgggttgcctc	21240
ggccgtggca gatgcgcgg tgcaggccgc cgcgtccggc gcttgcacca gctctacgg	21300
ggtgcaaaacg gacccgttggat ttttgcgtt ttcagccccc cggcgtccgc gccgttcaag	21360
gaagtacggc gccgcacgcg cgttacttgc cgaatatgtcc ctatcatcc ccatcgcc	21420
taccccccggc ttttgcgtt gatgttaccgc ccccaagaaga cggacttacca cccgacgcgg	21480
aaccaccact ggaacccggc gccgcgttgc cgcgtccggc cccgtgttcc ccccgatttc	21540
cgtgcgttcc gtttgcgttgcg aaggaggccg gacccgttgc cgcgttgcacc cgcgttacca	21600
ccccacgttcc gtttggccatcc cgggttgcgtt gtttgcgtt gatgttgcgc tcaacttgc	21660
cctccgtttc cccgttgcgcg gatgttgcgtt gatgttgcgtt gatgttgcgtt gatgttgc	21720
ccacggccctg acggggccggc tgcgttgcgtt gacccgttgc cggcggccgc cgttgcacc	21780
tcgcacgttgc ggcgttgcgtt ttttgcgtt gatgttgcgtt gatgttgcgtt gatgttgc	21840
cgttgcgttgc ggcgttgcgtt ttttgcgtt gatgttgcgtt gatgttgcgtt gatgttgc	21900
catgtggaaa aatcaaaata aatcaaaata aatcaaaata aatcaaaata aatcaaaata aatcaaaata	21960

tgttagaatg	gaagacatca	actttgcgtc	actggccccc	cgacacggct	cgcgcccggt	22020
catggaaac	tggcaagata	tcggcaccag	caatatgagc	ggtggcgcct	tcagctgggg	22080
ctcgctgtgg	agcggcatta	aaaatttcgg	ttccggcgtt	aagaactatg	gcagcaaagc	22140
ctggAACAGC	agcacaggcc	agatgcttag	ggacaagttg	aaagagcaaa	atttccaaca	22200
aaaggtgtta	gatggcctgg	cctctggcat	tagcggggt	gtggacctgg	ccaaaccaggc	22260
agtcaaaat	aagattaaca	gtaagcttg	tcccccct	cccgtagagg	agcctccacc	22320
ggccgtggag	acagtgtctc	cagaggggg	tggcggaaag	cgtccgcgac	ccgacaggga	22380
agaaaactctg	gtgacgcaaa	tagacgagcc	tccctcgat	gaggaggc	taaagcaagg	22440
cctccccacc	accggctccc	tcggcccat	ggctacccgg	gtgctggcc	agcacacacc	22500
cgtaacgcgt	gaccgtccctc	ccccggccga	caccggcag	aaacctgtgc	tgccagggtcc	22560
gtccggcgtt	gtttaacccc	gtcctagccg	cgcgtccctg	cggcgcgccc	ccaggggtcc	22620
gcatcggtt	cggcccgtag	ccagtgccaa	ctggcaaaagc	acactgaaca	gcacgtggg	22680
tttgggggt	caatccctga	agcggcga	atgctctga	tagctaactgt	gtcgatgtg	22740
tgtcatgtat	gctccatgt	cgccgcaga	ggagctgctg	agccgcgcg	cgccccgttt	22800
ccaagatgge	taccccttcg	atgatggcgc	agtggctta	catgcacatc	tcggggccagg	22860
acgcctcgga	gtacctgagc	ccccggctgg	tgcaggctcg	ccgcgcacc	gagacgtact	22920
tcagcctgaa	taacaaggtt	agaaaaccca	cggtggcgc	tacgcacgac	gtgaccacag	22980
accggctc	gctttgcgt	ctggcgttca	tccgggtgg	ccgcgaggat	actgcgtact	23040
cgtacaaggc	gccccgttcc	ctagctgtgg	gtgataaccc	tggtgtac	atggcttcca	23100
cgtacttga	catccggcgc	gtgtggaca	ggggccctac	tttaaagccc	tactctggca	23160
ctgcctacaa	cgcactggcc	ccccgggtt	cccccaactc	gtgcgagttg	gaacaaaatg	23220
aaactgcaca	agtggatgt	caagaactt	acgaagagga	aatatggac	aatgaagtc	23280
aggcgcgaga	acaggaacaa	gctaagaaaa	cccatgtata	tgcccaggct	ccactgtccg	23340
gaataaaaaat	aactaaagaa	ggtctacaaa	taggaactgc	cgacgcccaca	gtagcaggtg	23400
ccggcaaaaga	aattttcgca	gacaaaactt	ttcaacctga	accacaagta	ggagaatctc	23460
aatggaaacga	agcggatgccc	acagcagctg	gtggaaagggt	tcttaaaaag	acaactccca	23520
tggaaaccctg	ctatggctca	tacgctagac	ccaccaattc	caacggcgga	cagggcgta	23580
tgggtgaaca	aaatgttaaa	ttggaaaagtc	aagtgcgaat	gcaatttttt	tccacatcca	23640
caaatgcac	aaatgaagtt	aacaatata	acccaacagt	tgtattgtac	agcgaagatg	23700
taaaatcgga	aactccagat	atcatcttt	tttataaacc	taaaatgggg	gataaaaaatg	23760
ccaaagtct	gttggacaa	caagcaatgc	caaacagacc	aaattacatt	gttttagag	23820
acaattttat	tggcttcatg	tattacaaca	gcacaggtaa	catgggtgtc	cttgcgtggc	23880
aggcatcgca	gttgaacgct	gtttagatt	tgcaagacag	aaacacagag	ctgtcctacc	23940
agctttgtct	tgattcaatt	ggcgacagaa	caagatactt	ttcaatgtgg	aatcaagctg	24000
ttgacagcta	tgatccagat	gtcagaattt	ttgagaacca	tggaaactgag	gtgagttgc	24060
caaattattt	ctttcccttt	ggtggaaattt	ggattactga	cactttcaa	gtgtttaaaa	24120
caactgtgc	taacggggac	caaggcaata	ctacccggca	aaaagattca	acatttgcag	24180
aacgcata	aatagggggt	ggaaaataact	ttgcatgg	aattaacctg	aatgcacacc	24240
tatggaaaa	tttcctttac	tccaatattt	cgctgtaccc	ggcagacaag	ctaaaatata	24300
acccaccaa	tgtggaaata	tctgacaacc	ccaaacaccta	cgactacatc	aacaaggag	24360
tgtggctcc	tgggttctgt	gactgttaca	ttaaccttgg	ggcgcgctgg	tcttggact	24420
acatggacaa	cgttaatccc	ttaaccacc	accgcaatgc	gggcctgcgt	taccgcctca	24480
tgttgttgg	aaacggccgc	tacgtgcctt	ttcacattca	gtgtccccaa	aagtttttg	24540
ccattaaaa	cctccctctc	ctgcccaggct	catacacata	tgaatggaaac	ttcaggaagg	24600
atgttaacat	ggtctgcag	agctctctgg	gaaacgaccc	tagagttgac	ggggctagca	24660
ttaagtttga	cagcattttgt	ctttacgc	ccttcttccc	catggcccac	aacacggcct	24720
ccacgcttgg	agccatgctc	agaaaatgaca	ccaaacgacca	gtcctttaat	gactaccc	24780
ccggcccca	catgtatata	cccatacc	ccaaacgccc	caacgtgccc	atccatcc	24840
catcgccaa	ctggcagca	tttcgggtt	ggcccttcac	acgcttgaag	acaaaggaaa	24900
cccccttcc	gggatcagge	tgacccccc	actacaccta	ctctggctcc	ataccatacc	24960
ttgacggaaac	cttttatctt	aatcacac	ttaagaagg	ggccattact	tttactctt	25020
ctgttagctg	gccccggcaac	gaccgcctgc	ttactccaa	tgagttttag	attaagcgct	25080
cagttgacgg	ggagggctat	aacgttagctc	agtgcacat	gacaaaaggac	tggttcttag	25140
tgcagatgtt	ggccaactac	aatattggct	accagggtt	ctacattcca	gaaagctaca	25200
aagaccgcatt	gtactcg	ttcagaaaact	tccagcccat	gagccggca	gtgggtggac	25260

ataactaaata	caaagattat	cagcagggttgc	gaattatccat	ccagcataac	aactcaggct	25320
tcgttaggcta	cctcgctccc	accatgcgcg	agggacaagc	ttaccccgct	aatgttccct	25380
acccactaat	aggcaaaacc	gcccgttata	gtattaccca	aaaaaagttt	cttgcgacc	25440
gcacccctgtg	gcgcatacccc	ttctccagta	actttatgtc	catgggtgcg	ctcacagacc	25500
tggccaaaaa	ccttctctac	gcaaaactccg	cccacgcgc	agacatgacc	tttgagggtgg	25560
atcccataa	cgagcccacc	cttctttatg	ttttgttga	agtctttgac	gtggccgtg	25620
tgccaccagcc	gcaccgcggc	gtcategaga	ccgtgtacct	gcccacgccc	ttctcgcccg	25680
gcaacgcac	aacataaaga	agcaagcaac	atcaacaaca	gctgcccaca	tggctccag	25740
tgagcaggaa	ctgaaaagcca	ttgtcaaaaga	tctttgtgt	gggcccattt	ttttgggcac	25800
ctatgacaa	cgcttccccag	gctttgttcc	cccacacaag	ctgcgcgtcg	ccatagttaa	25860
cacggcccg	cgcgagactg	ggggcgtaca	ctggatggcc	tttgccttgg	acccgcgtc	25920
aaaaacatgc	taccccttttgc	agcccttgg	ctttctgtac	caacgtctca	agcagggtta	25980
ccagtttgc	tacgactac	tcctgcggcc	taggcatttgc	gccttctccc	ccgaccgctg	26040
tataacgcgt	gaaaagtcca	cccaaagcg	gcaggggccc	aactcgccg	cctgtggct	26100
attctgcgtc	atgtttctcc	acgccttgc	caactggccc	caaactccca	tggatcacaa	26160
ccccaccatg	aaccttatta	ccgggttacc	caactccatg	cttaacagtc	cccaggta	26220
gcccaccctg	cgcgcgaacc	aggaacagct	ctacagcttc	ctggagcgc	actcgcccta	26280
cttccgcage	cacagtgcgc	aaatttaggag	cgccacttct	ttttgtcact	tgaaaaacat	26340
gtaaaaataa	tgtacttagga	gacacttca	ataaaaggca	atgttttat	ttgtacactc	26400
tcgggtgatt	atttacccccc	acccttggcg	tctgcgcgt	ttaaaaatca	aagggttct	26460
gcccgcgtc	gctatgcgc	actggcaggg	acacgttgc	atactgggt	ttagtgtc	26520
actttaaactc	aggcacaacc	atccgcggca	gctcggtaa	tttttcactc	cacaggctgc	26580
gcaccatcac	caacgcgttt	agcaggctcg	gcgcgcgat	cttgaagtcg	cagggtggc	26640
ctccgcctg	cgcgcgcgag	ttgcgataca	cagggttaca	gcaactggaa	actatcagcg	26700
ccgggttgc	cacgcgtggc	agcacgtct	tgtcgaggat	catatccgcg	tccaggctc	26760
ccgcgttgc	caggcgaaac	ggaggtaact	ttgttagctg	ccttccaaaa	aagggtgc	26820
gcccaggctt	tgagttgcac	tcgcacccgt	gtggcatcag	aagggtaccg	tgcccagtct	26880
gggcgtttag	atacagcgcc	tgcatgaaag	ccttgcattgc	cttaaaagcc	acctgagcct	26940
ttgcgccttc	agagaagaac	atgcgcgaag	acttgcggga	aaactgattt	gcccgcacagg	27000
ccgcgtcatg	cacgcgcac	cttgcgtcgg	tgttggagat	ctgcaccaca	tttcggccccc	27060
accgttctt	cacgatcttgc	gccttgcatt	actgttcctt	cagcgcgc	tgcccgttt	27120
cgctcgta	atccatttca	atcacgtct	ccttatttat	cataatgtc	ccgtgttagac	27180
acttaagctc	gccttcgatc	tcagcgcgc	ggtcagcc	caacgcgc	cccgtggc	27240
cgtggtgctt	gtaggttacc	tctgcacaa	actgcaggta	cgcctgcagg	aatgc	27300
tcatcgta	aaaggcttt	ttgctgttgc	aggctcagtc	caacccgcgg	tgcttctcg	27360
ttagccagg	tttgcatac	gcccgcagag	cttccatgt	gtcaggcagt	agcttgaagt	27420
ttgcctttag	atcggttacc	acgtggtact	tgttccatcaa	cgcgccgc	gcctccatgc	27480
ccttctccca	cgcagacacg	atcggcaggc	tacgggttgc	tatcaccgt	ctttcacttt	27540
ccggttact	ggacttctcc	ttttccttgc	gcacccgc	accccgcc	actgggtcg	27600
cttcattcag	ccgcgcgcacc	gtgcgttac	ctcccttgc	gtgttgcatt	agcaccgggt	27660
ggttgtgaa	accaccatt	tgtagcgc	catcttctct	tttttccctg	ctgtccacga	27720
tcacctctgg	ggatggcg	cgctcggt	tggagaggg	gctgttctt	ttttttttgg	27780
acgcaatggc	caaatccgc	gtcgaggc	atggccgcgg	gctgggtgt	cgccgcacca	27840
gcccgcatt	tgacgagtct	tcttcgttgc	cgactcgc	acgcgc	gcgcgtttt	27900
ttggggcgc	gccccggaggc	ggcggcgc	gcgcacgggg	cgagacgtc	tccatgggt	27960
gtggacgtcg	cgccgcacc	cgtcgcgt	cggggtgt	ttcgcgc	tccttctccc	28020
gactggccat	ttcccttc	tataggcaga	aaaagatcat	ggagtca	gagaaggagg	28080
acagccta	cgccccctt	gagttgc	ccacccgc	caccgtgc	gccaacgcgc	28140
ctaccac	ccccgtcg	gcacccccc	ttgaggagg	ggaagtatt	atcgagcagg	28200
acccaggtt	tgtaa	gacgac	atcgct	accaacagag	gataaaaagc	28260
aagaccagg	cgacgc	gaaaacagg	aaaagtc	gccccggac	caaaggcatg	28320
gcgactac	atgttgg	gacgacgt	tgttgaag	tctgcgc	cagtgc	28380
ttatctgc	cgcg	gacgacgt	atgtgc	cgccatagc	gatgtc	28440
ttgcctac	acgc	ttctcacc	gcgtac	caa	acgc	28500
catgcgag	caacccgc	ctcaactt	acccgtatt	tgccgt	gagggt	28560

ccacccatca catcttttc caaaaactgca agataccct atccctgcgt gccaaccgca	28620
gccgagcgg acaagcagctg gccttgcggc agggcgctgt cataacctgt atccgcctcg	28680
tcgaccaagt gccaaaaatc tttgagggtc ttggacgcga cgagaagcgc gggcaaaacg	28740
ctctgcaaca aaaaaacacg gaaaatgaaa gtcactgtgg agtgctggg gaacttgagg	28800
gtgacaacgc gcccctagcc gtgctgaaac gcacatcgaa ggtcaccac tttgcctacc	28860
cgggacttaa cttacccccc aaggttatga gcacagtcat gagcggactg atcgtgcgc	28920
gtgcacgacc cctggagagg gatgaaact tgcaagaaca aaccggagg ggcctacccg	28980
cagttggcga tgagcagctg ggcgcgttgc ttgagacgcg cgagcctgca gacttgagg	29040
agcgacgcaa gctaattgtat ggcgcgtgc ttgttaccgt ggagcttgc tgcatgcgc	29100
ggttcttgc tgacccggag atgcagcga agctagagga aacgttgac tacaccttgc	29160
gccaggccta cgtgcgcagg gcctgcaaaa ttccaaacgt ggagcttgc aacctggct	29220
cctaccccttgg aattttgcac gaaaacccgc ttgggcaaaa cgtgttccat tccacgctca	29280
agggcggaggc gcccgcgcac tacgtcgcg actgcgttta cttatccatg tgctacaccc	29340
ggcaaaacggc catggggcgtg tggcagcgt gcctggagga ggcgcaccc aaggagctgc	29400
agaagctgtt aaagcaaaac ttgaaggacc tatggacggc cttcaacgg cgctccgtgg	29460
ccgcgcaccc ggcggacatt atttccccg aacgcgttgc taaaacccgt caacagggtc	29520
tgccagactt caccagtcaa agcatgttgc aaaaacttttag gaactttatc cttagcggtt	29580
caggaattct gccccccacc tgctgtgcgc ttccctagcga ctttgcggc attaagtacc	29640
gtgaatgcggc tccgcgcgtt tggggtcaact gtcaccccttgc gcaagcttgc aactaccc	29700
cctaccactc cgacatcatg gaagacgtga ggcgtgacgg cctactggg tgtaactgtc	29760
gctgcaacccat atgcaccccg caccgcctcc ttgtctgcaaa ttccacaactg cttagcgaaa	29820
gtcaattat cggtaacctt gagctgcagg gtccttcgccc tgacgaaaag tccgcggc	29880
cggggttggaa actcaactccg gggctgttgc cgtcggttta ctttcgcaaa tttgttaccc	29940
aggactacca cggccacggat attaggttct acgaagacca atccgcggc ccaaatgcgg	30000
agtttacccgc ctgcgttattt acccaggccg acatcccttgg ccaatttgc gccaattaaaca	30060
aagccggcca agagtttgcg ctacgaaagg gacgggggggtt ttacttgac ccccaactcc	30120
gcgaggagct caacccaaatc ccccccggc cgcagcccta tcagcaggcc cgccgccttg	30180
cttcccagga tggcacccaa aaagaagctg cagctgcgc cggccgcacc cacggacgg	30240
gaggaataact gggacagtca ggcagaggag gttttggacg aggaggagga gatgatggaa	30300
gactgggaca gcctagacga ggaagcttcc gaggccgaaag aggtgtcaga cgaacaccgg	30360
tcaccctcggt tcgcattttcc ctcggccggc ccccaagaaat cggcaaccgt tcccacgtt	30420
gctacaacctt ccgcgttccca ggcggccggc gcaactggccg ttgcggacc caaccgttgc	30480
tgggacacca ctggacccca ggcgggttaag tctaaggcgc cgccgcgtt agcccaagag	30540
caacaacacg gccaaggctt cccgcgttgc cgcgtgcaca agaaccatc agttgttgc	30600
ttgcaagact gggggggcaaa catctcccttgc cccggccgtt ttcttcttca ccatcacggc	30660
gtggccttcc cccgttaatcat cctgcatttac taccgttcatc tctacagccc ctactgcacc	30720
ggcggcagcg gcagcaacag cagccggccac gcagaagcaa aggccgcgg atagcaagac	30780
tctgacaaag cccaaagaaat ccacagccgc ggcagcagca ggaggaggag cactgcgtt	30840
ggcggccaaac gaacccgtat cgacccgcga gtttagaaac aggattttc ccactctgt	30900
tgctatattt caacagagca gggccaaaga acaagagctg aaaataaaaa acaggtcttct	30960
gcccgttcc accccgcagct gcctgtatca caaaaagcgaa gatcagttc ggcgcacgc	31020
ggaagacgcg gaggcttcttc tcaagcaata ctgcgcgtt actcttaagg actagttcg	31080
ccccccttcttcaaaatggaaact acgttcatc cagccgcac accccggc	31140
agccacgttc gtcagcgccca ttatggacaa ggaatattccc acgccttaca tttggagtt	31200
ccaggccacaa atggggcttgc cggctggagc tgcccaagac tactcaaccc gaataaaacta	31260
catgagcgcg ggaccccaaca tgatattttcc ggtcaacggaa atccgcggcc accggaaacgg	31320
aattttccctc gaacaggccg ctattaccac cacccctgt aataacccatc atcccgtag	31380
ttggcccgctt gcccgttaccaggaaatcc accactgttcc accactgtt tactccctt	31440
agacgccccag gcccggacttcc agatgacttcc ctcaggccgc cagcttgcgg gggcttcc	31500
tcacagggtt cggctggcccg ggcagggttacttcacccgtt aaaaatccagag ggcggaggat	31560
tcagctcaac gacgagtcgg tgagcttcc tcttggcttc cgtccggacg ggacatttca	31620
gatccggccgc gctggccgtt cttcatttac gcccgttgc ggcatacttca ctctgcagac	31680
ctcgccctcg gaggccgcgtt ccggaggcat tggaaacttca caatttatttgg aggagttcg	31740
gccttcgggtt tacttcacc cttttcttgg accccggcc cactaccgg accagtttat	31800
tcctcaactt gacgcccgtt gacactcgcc ggcggacttca gactgaatga ccactggaga	31860

ggcagagcaa ctgcgcctga	cacacctcga ccactgcccgc	cgccacaagt gctttggcccg	31920
cggctccggt gagttttgtt	actttgaatt gcccgaagag	catatcgagg gccccggcgca	31980
cggcgccgg ctcaccaccc	aggttagagct tacacgtac	ctgattcggg agttaccaa	32040
gcccggccctg ctatgtggagc	gggagcgggg tccctgtgtt	ctgaccgtgg tttgcaactg	32100
tcctaaccct ggattacatc	aagatcttat tccattcaac	taacaataaa cacacaataa	32160
attacttact taaaatcagt	cagcaaattct ttgtccagct	tatttcagcat caccccttt	32220
ccctcctccc aactctggta	tttcagcagc ctttagctg	cgaactttct ccaaagtcta	32280
aatgggatgt caaattcctc	atgttcttgc ccctccgcac	ccactatctt catattgttg	32340
agatgaaac gcccagacc	gtctgaagac accttcaacc	ctgtgtaccc atatgacacg	32400
gaaaccggcc ctccaactgt	gcctttcctt acccctccct	ttgtgtcgcc aaatgggttc	32460
caagaaagtc ccccccggagt	gctttcttg cgtcttttag	aacccttggt tacccacac	32520
ggcatgtttg cgctaaaaat	gggcagcggc ctgtccctgg	atcaggcagg caacccataa	32580
tcaaatacaa tcaactgtttc	tcaaccgtta aaaaaaaca	agtccaaat aactttggaa	32640
acatccgcgc cccttacagt	cagctcaggc gcccataacc	tggccacaac ttgcctttg	32700
gtggctctgt acaacactct	taccatgca tcacaaggcac	cgctaacccgt gcaagactca	32760
aaacttagca ttgttaccaa	agagccacatt agatgttag	atggaaaact ggccctgcag	32820
acatcaggccc ccctctctgc	cactgataaac aacgcctca	ctatcaactgc ctccacccct	32880
cttactactg caaatggtag	tctggctgtt accatggaaa	acccacttta caacaacaat	32940
ggaaaacttg ggctcaaaat	tggcggtctt ttgcaagtgg	ccaccgactc acatgcacta	33000
acacttaggtt ctggtcaggg	ggttgcagtt cataacaatt	tgctacatac aaaagttaca	33060
ggcgaatag ggtttgatac	atctggcaac atgaactta	aaactggaga tggctctat	33120
gtggatagcg ccgtcctaa	ccaaaaacta catattaatc	taaataccac aaaaggcctt	33180
gctttgaca acaccgcaat	aacaattaac gctggaaaag	ggttggaaatt tgaaacagac	33240
tcctcaaaacg gaaatcccatt	aaaaacaaaa attggatcag	gcataacaata taatccaat	33300
ggagctatgg ttgcaaaaact	tggAACAGC ctcagttttg	acagctcgg agccataaca	33360
atgggcagca taaacaatgaa	cagacttact ctttgacaa	caccagaccc atccccaaat	33420
tgcagaattt cttcagataa	agactgcaag ctaactctgg	cgctaacaaa atgtggcagt	33480
caatttttgg gcactgtttc	agctttggca gtatcaggta	atatggcctc catcaatgg	33540
actctaagca gtgtaaactt	ggttcttaga tttgtgaca	acggagtgt tatgtcaata	33600
tcatcaactgg acaaacagta	tttggacattt agaaacgggg	actccactaa cggtcaacca	33660
tacacttatg ctgttgggtt	tatgccaac ctaaaagctt	acccaaaaac tcaaagtaaa	33720
actgcaaaaa gtaatattgt	tagccaggtg tatcttaatg	gtgacaagtc taaaccattg	33780
cattttacta ttacgctaaa	tggAACAGC gaaaccaacc	aagtaagcaa atactcaata	33840
tcattcagtt ggtctctggaa	cagtggacaa tacactaatg	acaaatttgc caccaattcc	33900
tataccttct cctacattgc	ccaggaataa agaatcgtga	acctgttgc ttttatgttt	33960
caacgtgtt attttcaat	tgcagaaaat ttcaagtcat	ttttcatca gtagtatagc	34020
cccaccacca catagcttat	actaataccat gtaccttaat	caaactcaca gaaacccatgt	34080
attcaacactg ccacccctt	cccaacacac agatcaca	gtcctttctc cccggctggc	34140
cttaaacacg atcatatcat	gggttaacaga cataatttca	ggttattat tccacacgg	34200
ctccctgtcg gccaaacgct	catcagtgtt gtttataaaac	tccccggca gctcgcttaa	34260
gttcatgtcg ctgtccagct	gctgaggccac aggtctgtt	ccaaacttgcg gttgtcaac	34320
ggccggcgaa ggagaagtcc	acgcctacat ggggttagag	tcataatctt gcatcaggat	34380
aggggcgggtt tgctgcagca	gcccggcgaat aaactgtgc	cgccggccgt ccgtctgc	34440
ggaatacaac atggcagtgg	tcttcctcagc gatattcgc	accggccgca gcataaggcg	34500
ccttgcctc cgggcacagc	agcgcacccct gatctcaatt	aagtcaagcac agtaactgca	34560
gcacagtacc acaatattgt	ttaaaaatccc acagtgcag	gcccgttatac caaagctcat	34620
ggcggggacc acagaaccca	cgtggccatc ataccacaag	cgcaggtaga ttaagtggcg	34680
acccctcata aacacgctgg	acataaaacat taccttttt	ggcatgttgc aattcaccac	34740
ctcccggtac catataaacc	tctgattaaa catggcgcca	tccaccacca tcctaaacca	34800
gctggccaaa acctgcccgc	cggtatgc ctgcaggggaa	ccgggactgg aacaatgaca	34860
gtggagagcc caggactctgt	aaccatggat catcatgctc	gtcatgatcat caatgttgc	34920
acaacacagg cacacgtgca	tacacttctt caggattaca	agtcctccccc gctcagaac	34980
catatcccg ggaacaaccc	atccctgaat cagcgttaat	cccacactgc agggaaagacc	35040
tcgcacgtaa ctacgttgc	gcattgtcaa agtttacat	tcgggcagca gcggatgatc	35100
ctccagatgt gtagcgcggg	tttctgtctc aaaaggaggt	agacgatccc tactgtacgg	35160

agtgcgccga	gacaaccgag	atcgtgttgg	tcgttagtgtc	atgccaaatg	gaacgcccgg	35220
cgtagtcata	tttcctgaag	caaaaccagg	tgccggcgtg	acaaacagat	ctgcgtctcc	35280
ggtctcgccg	cttagatcgc	tctgtgtgt	agttgttagt	tatccactct	ctcaaagcat	35340
ccagggcccc	ccctggcttcg	ggttctatgt	aaactccttc	atgcgccgt	gccctgtataa	35400
catccaccac	cgcagaataa	gccacaccca	gccaacctac	acattcggtc	tgcgagtcac	35460
acacgggagg	agcgggaaga	gctggaaagaa	ccatgttttt	ttttttatcc	caaaagattta	35520
tccaaaacct	caaaatgaag	atctattaag	tgaacgcgt	ccccctccgt	ggcgtggtca	35580
aactctacag	ccaaagaaca	gataatggca	ttttaagat	gttgcacaat	ggcttccaaa	35640
aggcaaacgg	ccctcacgto	caagtggacg	taaaggctaa	acccttcagg	gtgaatctcc	35700
tctataaaca	ttccagcacc	ttcaaccatg	cccaaataat	tctcatctcg	ccaccccttc	35760
aatatatctc	taagcaaato	ccgaatattt	agttccggca	ttgtaaaaat	ctgctccaga	35820
gcccctcca	cccttcagect	caagcagcga	atcatgattt	caaaaattca	gttcttcac	35880
agacctgtat	aaggattcaaa	agcggaaacat	taacaaaaat	accgcgatcc	cgttaggtccc	35940
ttcgcaggc	cagctgaaca	taatcgtcga	ggttcgcacg	gaccagcgc	gccacttccc	36000
cgcaggaaac	catgacaaaa	gaacccacac	tgattatgac	acgcataactc	ggagctatgc	36060
taaccagct	agccccgatg	taagctgtt	gcatggcgg	cgatataaaa	tgcaagggtc	36120
tgctcaaaaa	atcaggcaaa	gcctcgca	aaaaagaaaag	cacatcgtag	tcatgctcat	36180
gcagataaaag	gcaggttaagc	tccggaaacca	ccacagaaaa	agacaccatt	tttctctcaa	36240
acatgtctgc	gggtttctgc	ataaaacacaa	aataaaataa	caaaaaaaaca	tttaaacatt	36300
agaaggctgt	cttacaacag	gaaaaacaac	ccttataagc	ataagacgga	ctacggccat	36360
gccggcgtga	ccgtaaaaaa	actggtacc	gtgataaaaa	agcaccaccc	acagctcctc	36420
ggtcatgtcc	ggagtcataa	tgtaagactc	ggtaaacaaca	tcaggttcat	tcacatcggt	36480
cagtctaaa	aagcgcacca	aatagcccg	gggaaatacat	acccgcaggc	gtagagacaa	36540
cattacagcc	cccatagggag	gtataacaaa	attaatagga	gagaaaaaca	cataaaccac	36600
tgaaaaaccc	tcctgcctag	gcaaaatagc	accctccgc	tccagaacaa	catacagcgc	36660
ttccacagcg	gcagccataa	cagtcagcct	taccagtaaa	aaagaaaaac	tattaaaaaa	36720
acaccactcg	acacggcacc	agctcaatca	gtcacagttt	aaaaaaggc	caagtgcaga	36780
gcgagttat	ataggactaa	aaaatgacgt	aacggttaaa	gtccacaaaa	aacacccaga	36840
aaacccgcacg	cgaacctacg	cccagaaacg	aaagccaaaa	aacccacaac	ttcctcaaat	36900
cgtcaattcc	gttttcccac	gttacgtcac	ttcccatttt	aagaaaacta	caattcccaa	36960
cacatacaag	ttactccgccc	ctaaaaccta	cgtcacccgc	cccgttccca	cgcggcgc	37020
cacgtcaca	actccaccccc	ctcattatca	tattggctt	aatccaaaat	aaggatattt	37080
attgatgtat						37090

<210> 5

<211> 5955

<212> DNA

<213> Artificial Sequence

<220>

<223> NS cDNA sequence

<221> CDS

<222> (1)...(5955)

<400> 5

atg	gcg	ccc	atc	acg	gcc	tac	tcc	caa	cag	acg	cgg	ggc	cta	ctt	ggt
Met	Ala	Pro	Ile	Thr	Ala	Tyr	Ser	Gln	Gln	Thr	Arg	Gly	Leu	Leu	Gly
1			5							10			15		

48

tgc	atc	atc	act	agc	ctt	aca	ggc	cgg	gac	aag	aac	cag	gtc	gag	gga
Cys	Ile	Ile	Thr	Ser	Leu	Thr	Gly	Arg	Asp	Lys	Asn	Gln	Val	Glu	Gly
20									25				30		

96

gag	gtt	cag	gtg	gtt	tcc	acc	gca	aca	caa	tcc	ttc	ctg	gcg	acc	tgc
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

144

Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys			
35	40	45	
gtc aac ggc gtg tgt tgg acc gtt tac cat ggt gct ggc tca aag acc			192
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr			
50	55	60	
tta gcc ggc cca aag ggg cca atc acc cag atg tac act aat gtg gac			240
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp			
65	70	75	80
cag gac ctc gtc ggc tgg cag gcg ccc cgg gcg cgt tcc ttg aca			288
Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr			
85	90	95	
cca tgc acc tgt ggc agc tca gac ctt tac ttg gtc acg aga cat gct			336
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala			
100	105	110	
gac gtc att ccg gtc cgc cgg cgg ggc gac agt agg ggg agc ctg ctc			384
Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu			
115	120	125	
tcc ccc agg cct gtc tcc tac ttg aag ggc tct tcg ggt ggt cca ctg			432
Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu			
130	135	140	
ctc tgc cct tcg ggg cac gct gtc ggc atc ttc cgg gct gcc gta tgc			480
Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys			
145	150	155	160
acc cgg ggg gtt gcg aag gcg gtc gac ttt gtc ccc gta gag tcc atg			528
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met			
165	170	175	
gaa act act atg cgg tct ccg gtc ttc acg gac aac tca tcc ccc cgg			576
Glü Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro			
180	185	190	
gcc gta ccg cag tca ttt caa gtc gcc cac cta cac gct ccc act ggc			624
Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly			
195	200	205	
agc ggc aag agt act aaa gtc ccg gct gca tat gca gcc caa ggg tac			672
Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr			
210	215	220	
aag gtc ctc ctc aat ccg tcc gtt gcc gct acc tta ggg ttt ggg			720
Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly			
225	230	235	240
gcg tat atg tct aag gca cac ggt att gac ccc aac atc aga act ggg			768
Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly			
245	250	255	

gta agg acc att acc aca ggc gcc ccc gtc aca tac tct acc tat ggc Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly	260	265	270	816	
aag ttt ctt gcc gat ggt ggt tgc tct ggg ggc gct tat gac atc ata Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Ala Tyr Asp Ile Ile	275	280	285	864	
ata tgt gat gag tgc cat tca act gac tcg act aca atc ttg ggc atc Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile	290	295	300	912	
ggc aca gtc ctg gac caa gcg gag acg gct gga gcg cgg ctt gtc gtg Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val	305	310	315	960	
ctc gcc acc gct acg cct ccg gga tcg gtc acc gtg cca cac cca aac Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn	325	330	335	1008	
atc gag gag gtg gcc ctg tct aat act gga gag atc ccc ttc tat ggc Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly	340	345	350	1056	
aaa gcc atc ccc att gaa gcc atc agg ggg gga agg cat ctc att ttc Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Arg His Leu Ile Phe	355	360	365	1104	
tgt cat tcc aag aag aag tgc gac gag ctc gcc gca aag ctg tca ggc Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly	370	375	380	1152	
ctc gga atc aac gct gtg gcg tat tac cgg ggg ctc gat gtg tcc gtc Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val	385	390	395	400	1200
ata cca act atc gga gac gtc gtt gtc gtg gca aca gac gct ctg atg Ile Pro Thr Ile Gly Asp Val Val Val Ala Thr Asp Ala Leu Met	405	410	415	1248	
acg ggc tat acg ggc gac ttt gac tca gtg atc gac tgt aac aca tgt Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys	420	425	430	1296	
gtc acc cag aca gtc gac ttc acg ttg gat ccc acc ttc acc att gag Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu	435	440	445	1344	
acg acg acc gtg cct caa gac gca gtg tcg cgc tcg cag cgg cgg ggt Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly	450	455	460	1392	
agg act ggc agg ggt agg aga ggc atc tac agg ttt gtg act ccg gga Arg Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly	465	470	475	480	1440

gaa cgg ccc tcg ggc atg ttc gat tcc tcg gtc ctg tgt gag tgc tat Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr 485 490 495	1488
gac gcg ggc tgg gct tgg tac gag ctc acc ccc gcc gag acc tcg gtt Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val 500 505 510	1536
agg ttg cgg gcc tac ctg aac aca cca ggg ttg ccc gtt tgc cag gac Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp 515 520 525	1584
cac ctg gag ttc tgg gag agt gtc ttc aca ggc ctc acc cac ata gat His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp 530 535 540	1632
gca cac ttc ttg tcc cag acc aag cag gca gga gac aac ttc ccc tac Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr 545 550 555 560	1680
ctg gta gca tac caa gcc acg gtg tgc gcc agg gct cag gcc cca cct Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro 565 570 575	1728
cca tca tgg gat caa atg tgg aag tgt ctc ata cgg ctg aaa cct acg Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Pro Thr 580 585 590	1776
ctg cac ggg cca aca ccc ttg ctg tac agg ctg gga gcc gtc caa aat Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn 595 600 605	1824
gag gtc acc ctc acc cac ccc ata acc aaa tac atc atg gca tgc atg Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met 610 615 620	1872
tcg gct gac ctg gag gtc act agc acc tgg gtg ctg gtg ggc gga Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly 625 630 635 640	1920
gtc ctt gca gct ctg gcc gcg tat tgc ctg aca aca ggc agt gtg gtc Val Leu Ala Ala Leu Ala Tyr Cys Leu Thr Thr Gly Ser Val Val 645 650 655	1968
att gtg ggt agg att atc ttg tcc ggg agg ccg gct att gtt ccc gac Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp 660 665 670	2016
agg gag ttt ctc tac cag gag ttc gat gaa atg gaa gag tgc gcc tcg Arg Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser 675 680 685	2064
cac ctc cct tac atc gag cag gga atg cag ctc gcc gag caa ttc aag	2112

His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys			
690	695	700	
cag aaa gcg ctc ggg tta ctg caa aca gcc acc aaa caa gcg gag gct			2160
Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala			
705	710	715	720
gct gct ccc gtg gtg gag tcc aag tgg cga gcc ctt gag aca ttc tgg			2208
Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp			
725	730	735	
gcg aag cac atg tgg aat ttc atc aac ggg ata cag tac tta gca ggc			2256
Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly			
740	745	750	
tta tcc act ctg cct ggg aac ccc gca ata gca tca ttg atg gca ttc			2304
Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe			
755	760	765	
aca gcc tct atc acc agc ccg ctc acc acc caa agt acc ctc ctg ttt			2352
Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe			
770	775	780	
aac atc ttg ggg ggg tgg gtg gct gcc caa ctc gcc ccc ccc agc gcc			2400
Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala			
785	790	795	800
gct tcg gct ttc gtg ggc gcc atc gcc ggt gct gtt ggc agc			2448
Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser			
805	810	815	
ata ggc ctt ggg aag gtg ctt gtg gac att ctg gcg ggt tat gga gca			2496
Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala			
820	825	830	
gga gtg gcc ggc gcg ctc gtg gcc ttc aag gtc atg agc ggc gag atg			2544
Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met			
835	840	845	
ccc tcc acc gag gac ctg gtc aat cta ctt cct gcc atc ctc tct cct			2592
Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro			
850	855	860	
ggc gcc ctg gtc gtc ggg gtc gtg tgt gca gca ata ctg cgt cga cac			2640
Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His			
865	870	875	880
gtg ggt ccg gga gag ggg gct gtg cag tgg atg aac cgg ctg ata gcg			2688
Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala			
885	890	895	
ttc gcc tcg cgg ggt aat cat gtt tcc ccc acg cac tat gtg cct gag			2736
Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu			
900	905	910	

agc gac gcc gca gcg cgt gtt act cag atc ctc tcc agc ctt acc atc Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile 915 920 925	2784
act cag ctg ctg aaa agg ctc cac cag tgg att aat gaa gac tgc tcc Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser 930 935 940	2832
aca ccg tgt tcc ggc tcg tgg cta agg gat gtt tgg gac tgg ata tgc Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys 945 950 955 960	2880
acg gtg ttg act gac ttc aag acc tgg ctc cag tcc aag ctc ctg ccg Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro 965 970 975	2928
cag cta ccg gga gtc cct ttt ttc tcg tgc caa cgc ggg tac aag gga Gln Leu Pro Gly Val Pro Phe Ser Cys Gln Arg Gly Tyr Lys Gly 980 985 990	2976
gtc tgg cgg gga gac ggc atc atg caa acc acc tgc cca tgt gga gca Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala 995 1000 1005	3024
cag atc acc gga cat gtc aaa aac ggt tcc atg agg atc gtc ggg cct Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro 1010 1015 1020	3072
aag acc tgc agc aac acg tgg cat gga aca ttc ccc atc aac gca tac Lys Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr 1025 1030 1035 1040	3120
acc acg ggc ccc tgc aca ccc tct cca gcg cca aac tat tct agg gcg Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala 1045 1050 1055	3168
ctg tgg cgg gtg gcc gct gag gag tac gtg gag gtc acg cgg gtg ggg Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly 1060 1065 1070	3216
gat ttc cac tac gtg acg ggc atg acc act gac aac gta aag tgc cca Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro 1075 1080 1085	3264
tgc cag gtt ccg gct cct gaa ttc acg gag gtg gac gga gtg cgg Cys Gln Val Pro Ala Pro Glu Phe Thr Glu Val Asp Gly Val Arg 1090 1095 1100	3312
ttg cac agg tac gct ccg gcg tgc agg cct ctc cta cgg gag gag gtt Leu His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val 1105 1110 1115 1120	3360
aca ttc cag gtc ggg ctc aac caa tac ctg gtt ggg tca cag cta cca Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro 1125 1130 1135	3408

tgc gag ccc gaa ccg gat gta gca gtg ctc act tcc atg ctc acc gac Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp 1140 1145 1150	3456
ccc tcc cac atc aca gca gaa acg gct aag cgt agg ttg gcc agg ggg Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly 1155 1160 1165	3504
tct ccc ccc tcc ttg gcc agc tct tca gct agc cag ttg tct gcg cct Ser Pro Pro Ser Leu Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro 1170 1175 1180	3552
tcc ttg aag gcg aca tgc act acc cac cat gtc tct ccg gac gct gac Ser Leu Lys Ala Thr Cys Thr His His Val Ser Pro Asp Ala Asp 1185 1190 1195 1200	3600
ctc atc gag gcc aac ctc ctg tgg cgg cag gag atg ggc ggg aac atc Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile 1205 1210 1215	3648
acc cgc gtg gag tcg gag aac aag gtg gta gtc ctg gac tct ttc gac Thr Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp 1220 1225 1230	3696
ccg ctt cga gcg gag gag gat gag agg gaa gta tcc gtt ccg gcg gag Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu 1235 1240 1245	3744
atc ctg cgg aaa tcc aag aag ttc ccc gca gcg atg ccc atc tgg gcg Ile Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala 1250 1255 1260	3792
cgc ccg gat tac aac cct cca ctg tta gag tcc tgg aag gac ccg gac Arg Pro Asp Tyr Asn Pro Pro Leu Leu Ser Trp Lys Asp Pro Asp 1265 1270 1275 1280	3840
tac gtc cct ccg gtg gtg cac ggg tgc ccg ttg cca cct atc aag gcc Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala 1285 1290 1295	3888
cct cca ata cca cct cca ccg aga aag agg acg gtt gtc cta aca gag Pro Pro Ile Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu 1300 1305 1310	3936
tcc tcc gtg tct tct gcc tta gcg gag ctc gct act aag acc ttc ggc Ser Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly 1315 1320 1325	3984
agc tcc gaa tca tcg gcc gtc gac agc ggc acg gcg acc gcc ctt cct Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro 1330 1335 1340	4032
gac cag gcc tcc gac gac ggt gac aaa gga tcc gac gtt gag tcg tac	4080

Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr
 1345 1350 1355 1360
 tcc tcc atg ccc ccc ctt gag ggg gaa ccg ggg gac ccc gat ctc agt 4128
 Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser
 1365 1370 1375
 gac ggg tct tgg tct acc gtg agc gag gaa gct agt gag gat gtc gtc 4176
 Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val
 1380 1385 1390
 tgc tgc tca atg tcc tac aca tgg aca ggc gcc ttg atc acg cca tgc 4224
 Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys
 1395 1400 1405
 gct gcg gag gaa agc aag ctg ccc atc aac gcg ttg agc aac tct ttg 4272
 Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu
 1410 1415 1420
 ctg cgc cac cat aac atg gtt tat gcc aca aca tct cgc agc gca ggc 4320
 Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly
 1425 1430 1435 1440
 ctg cgg cag aag gtc acc ttt gac aga ctg caa gtc ctg gac gac 4368
 Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp
 1445 1450 1455
 cac tac cgg gac gtg ctc aag gag atg aag gcg aag gcg tcc aca gtt 4416
 His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val
 1460 1465 1470
 aag gct aaa ctc cta tcc gta gag gaa gcc tgc aag ctg acg ccc cca 4464
 Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro
 1475 1480 1485
 cat tcg gcc aaa tcc aag ttt ggc tat ggg gca aag gac gtc cgg aac 4512
 His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn
 1490 1495 1500
 cta tcc agc aag gcc gtt aac cac atc cac tcc gtg tgg aag gac ttg 4560
 Leu Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu
 1505 1510 1515 1520
 ctg gaa gac act gtg aca cca att gac acc acc atc atg gca aaa aat 4608
 Leu Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn
 1525 1530 1535
 gag gtt ttc tgt gtc caa cca gag aaa gga ggc cgt aag cca gcc cgc 4656
 Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg
 1540 1545 1550
 ctt atc gta ttc cca gat ctg gga gtc cgt gta tgc gag aag atg gcc 4704
 Leu Ile Val Phe Pro Asp Leu Glu Val Arg Val Cys Glu Lys Met Ala
 1555 1560 1565

ctc tat gat gtg gtc tcc acc ctt cct cag gtc gtg atg ggc tcc tca Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser 1570 1575 1580	4752
tac gga ttc cag tac tct cct ggg cag cga gtc gag ttc ctg gtg aat Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn 1585 1590 1595 1600	4800
acc tgg aaa tca aag aaa aac ccc atg ggc ttt tca tat gac act cgc Thr Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg 1605 1610 1615	4848
tgt ttc gac tca acg gtc acc gag aac gac atc cgt gtt gag gag tca Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser 1620 1625 1630	4896
att tac caa tgt tgt gac ttg gcc ccc gaa gcc aga cag gcc ata aaa Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys 1635 1640 1645	4944
tcg ctc aca gag cgg ctt tat atc ggg ggt cct ctg act aat tca aaa Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys 1650 1655 1660	4992
ggg cag aac tgc ggt tat cgc cgg tgc cgc gcg agc ggc gtg ctg acg Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr 1665 1670 1675 1680	5040
act agc tgc ggt aac acc ctc aca tgt tac ttg aag gcc tct gca gcc Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala 1685 1690 1695	5088
tgt cga gct gcg aag ctc cag gac tgc acg atg ctc gtg aac gga gac Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp 1700 1705 1710	5136
gac ctt gtc gtt atc tgt gaa agc gcg gga acc caa gag gac gcg gcg Asp Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala 1715 1720 1725	5184
agc cta cga gtc ttc acg gag gct atg act agg tac tct gcc ccc ccc Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro 1730 1735 1740	5232
ggg gac ccg ccc caa cca gaa tac gac ttg gag ctg ata aca tca tgt Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys 1745 1750 1755 1760	5280
tcc tcc aat gtg tcg gtc gcc cac gat gca tca ggc aaa agg gtg tac Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr 1765 1770 1775	5328
tac ctc acc cgt gat ccc acc acc ccc ctc gca cgg gct gcg tgg gaa Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu 1780 1785 1790	5376

aca gct aga cac act cca gtt aac tcc tgg cta ggc aac att atc atg Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met 1795 1800 1805	5424
tat gcg ccc act ttg tgg gca agg atg att ctg atg act cac ttc ttc Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe 1810 1815 1820	5472
tcc atc ctt cta gca cag gag caa ctt gaa aaa gcc ctg gac tgc cag Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln 1825 1830 1835 1840	5520
atc tac ggg gcc tgg tac tcc att gag cca ctt gac cta cct cag atc Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile 1845 1850 1855	5568
att gaa cga ctc cat ggc ctt agc gca ttt tca ctc cat agt tac tct Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser 1860 1865 1870	5616
cca ggt gag atc aat agg gtg gct tca tgc ctc agg aaa ctt ggg gta Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val 1875 1880 1885	5664
cca ccc ttg cga gtc tgg aga cat cgg gcc agg agc gtc cgc gct agg Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg 1890 1895 1900	5712
cta ctg tcc cag ggg ggg agg gcc gac act tgg ggc aag tac ctc ttc Leu Leu Ser Gln Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe 1905 1910 1915 1920	5760
aac tgg gca gtg aag acc aaa ctc aaa ctc act cca atc ccg gct gcg Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala 1925 1930 1935	5808
tcc cag ctg gac ttg tcc ggc tgg ttc gtt gct ggt tac agc ggg gga Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly 1940 1945 1950	5856
gac ata tat cac agc ctg tct cgt gcc cga ccc cgc tgg ttc atg ctg Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu 1955 1960 1965	5904
tgc cta ctc cta ctt tct gta ggg gta ggc atc tac ctg ctc ccc aac Cys Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn 1970 1975 1980	5952
cga Arg 1985	5955
<210> 6 <211> 1984	

<212> PRT
 <213> Artificial Sequence

<220>
 <223> NS sequence

<400> 6
 Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
 1 5 10 15
 Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu
 20 25 30
 Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
 35 40 45
 Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu
 50 55 60
 Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln
 65 70 75 80
 Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro
 85 90 95
 Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
 100 105 110
 Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
 115 120 125
 Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
 130 135 140
 Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr
 145 150 155 160
 Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu
 165 170 175
 Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
 180 185 190
 Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
 195 200 205
 Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
 210 215 220
 Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
 225 230 235 240
 Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val
 245 250 255
 Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys
 260 265 270
 Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
 275 280 285
 Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly
 290 295 300
 Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
 305 310 315 320
 Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile
 325 330 335
 Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
 340 345 350
 Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys
 355 360 365
 His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu
 370 375 380

Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
 385 390 395 400
 Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
 405 410 415
 Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
 420 425 430
 Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
 435 440 445
 Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg
 450 455 460
 Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
 465 470 475 480
 Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
 485 490 495
 Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg
 500 505 510
 Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
 515 520 525
 Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
 530 535 540
 His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
 545 550 555 560
 Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
 565 570 575
 Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
 580 585 590
 His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
 595 600 605
 Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser
 610 615 620
 Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val
 625 630 635 640
 Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile
 645 650 655
 Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp Arg
 660 665 670
 Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His
 675 680 685
 Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln
 690 695 700
 Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala
 705 710 715 720
 Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp Ala
 725 730 735
 Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu
 740 745 750
 Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr
 755 760 765
 Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe Asn
 770 775 780
 Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala
 785 790 795 800
 Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser Ile
 805 810 815
 Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly

820	825	830	
Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met Pro			
835	840	845	
Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly			
850	855	860	
Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val			
865	870	875	880
Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe			
885	890	895	
Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser			
900	905	910	
Asp Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile Thr			
915	920	925	
Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser Thr			
930	935	940	
Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys Thr			
945	950	955	960
Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro Gln			
965	970	975	
Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly Val			
980	985	990	
Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala Gln			
995	1000	1005	
Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro Lys			
1010	1015	1020	
Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr Thr			
1025	1030	1035	1040
Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala Leu			
1045	1050	1055	
Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly Asp			
1060	1065	1070	
Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro Cys			
1075	1080	1085	
Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg Leu			
1090	1095	1100	
His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val Thr			
1105	1110	1115	1120
Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro Cys			
1125	1130	1135	
Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro			
1140	1145	1150	
Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly Ser			
1155	1160	1165	
Pro Pro Ser Leu Ala Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser			
1170	1175	1180	
Leu Lys Ala Thr Cys Thr Thr His His Val Ser Pro Asp Ala Asp Leu			
1185	1190	1195	1200
Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr			
1205	1210	1215	
Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp Pro			
1220	1225	1230	
Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu Ile			
1235	1240	1245	
Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala Arg			
1250	1255	1260	

Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp Tyr
 1265 1270 1275 1280
 Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala Pro
 1285 1290 1295
 Pro Ile Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu Ser
 1300 1305 1310
 Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly Ser
 1315 1320 1325
 Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro Asp
 1330 1335 1340
 Gin Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr Ser
 1345 1350 1355 1360
 Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp
 1365 1370 1375
 Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val Cys
 1380 1385 1390
 Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala
 1395 1400 1405
 Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu
 1410 1415 1420
 Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly Leu
 1425 1430 1435 1440
 Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His
 1445 1450 1455
 Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys
 1460 1465 1470
 Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His
 1475 1480 1485
 Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu
 1490 1495 1500
 Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu Leu
 1505 1510 1515 1520
 Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu
 1525 1530 1535
 Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu
 1540 1545 1550
 Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu
 1555 1560 1565
 Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser Tyr
 1570 1575 1580
 Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr
 1585 1590 1595 1600
 Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys
 1605 1610 1615
 Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile
 1620 1625 1630
 Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser
 1635 1640 1645
 Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly
 1650 1655 1660
 Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr
 1665 1670 1675 1680
 Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys
 1685 1690 1695

Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp Asp
 1700 1705 1710
 Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser
 1715 1720 1725
 Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly
 1730 1735 1740
 Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser
 1745 1750 1755 1760
 Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr
 1765 1770 1775
 Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr
 1780 1785 1790
 Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr
 1795 1800 1805
 Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser
 1810 1815 1820
 Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile
 1825 1830 1835 1840
 Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile
 1845 1850 1855
 Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro
 1860 1865 1870
 Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro
 1875 1880 1885
 Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu
 1890 1895 1900
 Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn
 1905 1910 1915 1920
 Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser
 1925 1930 1935
 Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp
 1940 1945 1950
 Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu Cys
 1955 1960 1965
 Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1970 1975 1980

<210> 7

<211> 4909

<212> DNA

<213> Artificial Sequence

<220>

<223> pV1J nucleic acid

<400> 7

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcggttg	120
ttggcggtg	tcggggctgg	cttaactatg	cgccatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgttaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataatatg	tacatttata	ttggctcatg	300
tccaaacatta	ccgcatatgtt	gacatttgatt	attgactagt	tatataatgt	aatcaattac	360
gggtcatta	gttcatagcc	cataatatgg	gttccgcgtt	acataactta	cggtaaatgg	420
ccgcctggc	tgaccgcccc	acgaccggcc	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggagtatt	tacggtaaac	540

tgcccacccat	gcagttacatc	aagtgttatca	tatgccaagt	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggccgcct	ggcattatgc	ccagttacatg	accttatggg	actttctac	660
ttggcagtagc	atctacgtat	tagtcatcgc	tattaccatg	gtatgcgtt	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgcgtc	acggggattt	ccaagtctcc	acccattga	780
cgtcaatggg	agtttgcgtt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaccaa	840
ctccgcccc	ttgacgaaa	tggcggtag	gctgtacgg	tggaggctt	atataaggcag	900
agctcgat	gtgaaccgtc	agatcgctg	gagacgccc	ccacgctgtt	ttgacctcca	960
tagaagacac	cggaccgtat	ccagcctccg	cggccggaa	cggtgcattt	gaacgcggat	1020
ccccctgtcc	aagagtgacg	taagtaccgc	ctatagactc	tataggcaca	cccccttggc	1080
tcttatgcat	gtatatactgt	tttggcttg	gggcctatac	accccccgtt	ccttatgtca	1140
taggtatgg	tatagcttag	cctataggt	tgggttattt	accattattt	accatcccc	1200
tatgggtac	gatactttcc	attactaatac	cataacatgg	ctcttgcac	caactatctc	1260
tatggctat	atgccaatac	tctgtccctt	agagactgac	acggacttgc	tatttttaca	1320
ggatgggtc	ccatttata	tttacaattt	ccatataca	acaacgcctg	ccccctgtcc	1380
cgcagtttt	attaaacata	gctgtggatc	tccacgcga	tctcgggtac	gtgtccgg	1440
catgggtct	tctccgtat	cggcgagct	tccacatccg	agccctggtc	ccatgcctcc	1500
agcggctcat	ggtcgctcg	cagctccctt	ctccctaacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtggcg	cacaaggccg	tggcggtagg	gtatgtgtct	1620
aaaaatgagc	gtggagattt	ggctcgac	gctgacgcag	atggaaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagttt	gtattctgtat	aagagtca	ggtactccc	1740
gttgcggtgc	tgttaacggt	ggagggcagt	gtagtcgtat	cagtactcgt	tgctgcgcg	1800
cgcgccacca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggtcttttgc	1860
tgcaagtca	gtcccttagat	ctaggtacca	gatatcagaa	ttcagtcac	agcggcccg	1920
atctgctgt	cctcttagtt	gccagccatc	tgtgtttgc	ccctccccc	tgccttcctt	1980
gaccctggaa	ggtgcactc	ccactgtctt	ttcctaataa	aataggaaaa	ttgcatcgca	2040
ttgtctgt	agggttgcatt	ctattctgg	gggtgggggt	gggcaggac	gcaaaaaaaaa	2100
ggattggaa	gacaatagca	ggcatgttgg	ggatgcgtt	ggctctatgg	ccgctgcggc	2160
cagggtctga	agaattgacc	cggttccctt	tggccagaa	agaagcaggc	acatcccc	2220
ctctgtgaca	caccctgtcc	acgccccctt	tttttagttc	cagccccact	cataggacac	2280
tcatacgat	ggagggctcc	gccttcaatc	ccacccgcta	aagtacttgg	agcggctctc	2340
ccctccctca	tcaagcccacc	aaaccaaaacc	tagcctccaa	gagtggaaag	aaattaaagc	2400
aagataggct	atthaatgtca	gagggagaga	aaatgcctcc	aacatgttgg	gaagtaatga	2460
gagaatcat	agaatttctt	ccgcttc	gctcaactgac	tcgctgcgt	cggtcggtcg	2520
gctcggcga	gcccgtatcg	ctcaactcaaa	ggcggtataa	cggttatcc	cagaatcagg	2580
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagca	aaggccagga	accgtaaaaa	2640
ggccgcgtt	ctggcgtttt	tccataggt	ccgcccccc	gacgacatc	acaaaaaatcg	2700
acgctcaagt	cagagggtgc	gaaaccccgac	aggactataa	agataccagg	cgttcccc	2760
tggaaaggat	ctcggtcgct	ctctgttcc	gaccctgcgg	cttaccggat	acctgtccgc	2820
ctttccctt	tcgggaaaggcg	tggcgctt	tcatacgatca	cgctgttaggt	atctcgttcc	2880
ggtgttaggtc	gttcgctcca	agctggctg	tgtgcacgaa	ccccccgttc	agcccgaccg	2940
ctgcgcctt	tccggtaact	atcgcttgc	gtcccaaccgg	gtaagacacg	acttatcgcc	3000
actggcagca	gccactggta	acaggattag	cagagcggagg	tatgttaggc	gtgtacaga	3060
gttcttgaag	tggtggccta	actacggcta	cactagaaga	acagtatttgc	gtatctgcgc	3120
tctgtgt	ccagttacat	tcggaaaaag	agttggtagc	tcttgatcc	gcaaaacaaac	3180
caccgcttgt	agccgtgggt	tttttgtt	caagcagcag	attacgcga	aaaaaaaagg	3240
atctcaagaa	gatcccttgc	tctttctac	gggtctgcac	gctcagtgg	acgaaaaactc	3300
acgttaaggg	attttgcata	tgagattatc	aaaaaggatc	ttcacctga	tcctttaaa	3360
ttaaaaatga	agttttaat	caatctaaag	tatatatgag	taaacttggt	ctgacagttt	3420
ccaatgttca	atcagtgtt	cacccatctc	agcgatctgt	ctatttcggt	catccatagt	3480
tgccgtactc	gggggggggg	ggcgtgttgg	tctgcctgt	gaagaaggtt	ttgtgtactc	3540
ataccaggcc	tgaatcgccc	catatccag	ccagaaatgt	agggagccac	ggttgtatgag	3600
agcttgcgtt	taggtggacc	agttggtagt	tttgcacttt	tgctttggca	cggaacgcgtc	3660
tgcgttgcgt	gaaagatgcg	tgatctgtatc	tttgcactca	gcaaaaagttc	gatttattca	3720
acaaaggccgc	cgtccccgtca	agtcagcgat	atgtctgc	agtgttacaa	ccaattaacc	3780
aattctgtt	agaaaaactc	atcgagcatc	aaatgaaact	gcaattttt	catatcagga	3840

ttatcaatac catatTTTT	aaaaaggcgt ttctgtaatg	aaggagaaaa ctcaccgagg	3900
cagttccata ggatggcaag	atcctgtat cggtctgcga	ttccgactcg tccaaacatca	3960
ataacaacctttaatttccc	ctcgtcaaaa ataaggttat	caagtggaaa atcaccatga	4020
gtgacgactg aatccgggtg	gaatggcaaa agtttatgca	tttcttccca gacttgttca	4080
acaggccagc cattacgctc	gtcatcaaaa tcactcgcat	caacccaaacc gttattcatt	4140
cgtgattgcg cctgagcgag	acgaaatacg cgatcgctgt	taaaaggaca attacaaaca	4200
ggaatcaaatgcaaccggcg	caggaacact gccagcgcat	caacaatatt ttcacctgaa	4260
tcaggatatt cttctaatac	ctggaatgtctttcccg	ggatcgcaat ggtgagtaac	4320
catgcatcat caggagtagc	gataaaaatgc ttgatggtc	gaagaggcat aaattccgtc	4380
agccagtttgcgtaccat	ctcatctgta acatcattgg	caacgctacc tttgccatgt	4440
ttcagaaaca actctggcgc	atcgggttccatacaatac	gatagattgt cgacccctgt	4500
tgcccacat tatecgccg	ccatttataccatataaaat	cagcatccat gttgaaattt	4560
aatcgccggcc tggagcaaga	cgtttccgt tgaatatggc	tcataacacc ccttgttatta	4620
ctgttatgtt aacgacagac	ttttattgtt catatgata	tatttttatac ttgtgcaatg	4680
taacatcaga gattttgaga	cacaacgtgg ctttcccccc	ccccccatata ttaaagcatt	4740
tatcaggggtt attgtctcat	gagcgatatac atatttgaat	gtattttagaa aaataaaca	4800
ataggggttc cgccacatt	tcccccgggggtt	acgtctaaga aaccattatt	4860
atcatgacat taacctataa	aaataggcgt atcacgaggc	cctttcgctc	4909

<210> 8

<211> 35935

<212> DNA

<213> Adenovirus serotype 6

<400> 8

catcatcaat aataatacctt	atttggatt gaagccaata	tgataatgag ggggtggagt	60
ttgtgacgtg ggcggggcg	tgggaacggg gcggtgacg	tagtagtgtg gcggaagtgt	120
gatgttgc当地	gtgtggcga acacatgtaa	gcgacggatg	180
gtgtgc当地	gtgtacacag gaagtggacaa	tttgc当地	240
taaatttggg	cgttaaccgag taagattgg	ccatttgc当地	300
agtgaatct	gataataattt gtgttactca	tagcgctaa	360
gactttgacc	gtttacgttg agactcgccc	agggtttt	420
cgggtcaaaag	ttggcgtttt attattatag	tcagctgacg	480
tgagttcctc	aaagaggccac tcttggatgc	cagcgatag	540
tccgacaccg	ggactgaaaa tgagacat	tatcgccac	600
aatggccgccc	agtctttgg accagctgtat	cgaagaggtt	660
tcctagccat	tttgaaccac ctacccttca	cgaactgtat	720
cgaagatccc	aacgaggagg cggtttgc当地	gattttccc	780
gcaggaaggg	attgacttac tcactttcc	gcccgc当地	840
ccttcccg	cagccccgagc agccggagca	gagagcctt	900
ccttgc当地	gagggtatcg atcttactt	ccacggaggct	960
cgaggatgaa	gagggtgagg agtttggat	agatattgt	1020
caggcttgc当地	cattatcacc ggaggaatac	gggggaccca	1080
ctatatgagg	acctgtggca tttttgtctt	cagtagtgc当地	1140
tagatggtg	ggtttggat ggtatattttt	tttttaattt	1200
gaattttgtat	tttgatattt tttaaaagg	cctgtgtctt	1260
ccagaaccgg	accttgc当地 acctaccgc	cgttctt当地	1320
cggccgacat	cacccgttgc tagagaatgc	aatatgtt	1380
ccttctaaca	cacccgttgc gatacacc	cgtgggtt	1440
gccgtgagag	gtgttccgc tcgcccaggct	tttttttgc当地	1500
cctggccaa	cttggactt gagctgtaaa	gtggatgt	1560
ttgc当地	gtttaacgc tttgttgc当地	aatatgtt	1620
gagataatgt	ttaacttgc当地 tggctgtt	aataaaagggt	1680
cggccgtggc	aatggggcg	ggcttaaagg gtatataatg	1740
ttttctgtctt	taatcttgc当地 tacatctgac	cttgggatgt tttggaaagat	1800
tgctgtactt	gctggaaacag agctctaaca	gtaccccttgc gttttggagg	

tttctgtggg gctcatccca	ggcaaaggta	gtctgcagaa	ttaaggagga	ttacaagtgg	1860
gaatttgaag agctttgaa	atccctgtgt	gagctgtttg	attctttgaa	tctgggtcac	1920
caggcgctt tccaagagaa	ggtcataag	actttggatt	tttccacacc	ggggcgcgct	1980
gcggctgctg ttgtttttt	gagtttata	aaggataaaat	ggagcgaaga	aaccatctg	2040
agcgggggtt acctgctgga	tttctggcc	atgcatactgt	ggagagcggt	tgtgagacac	2100
aagaatcgcc tgetactgtt	gtcttccgtc	cgcccgccga	taataccgac	ggaggaggcag	2160
cagcagcagc aggaggaagc	caggcgggg	cgccaggagc	agagcccatg	gaaccgcaga	2220
gccggcctgg accctcgga	atgaatgtt	tacaggtggc	tgaactgtat	ccagaactga	2280
gacgcattt gacaattaca	gaggatggc	agggctaaa	ggggtaaag	agggagcggg	2340
gggcttgtga ggctacagag	gaggctagga	atctagctt	tagcttaatg	accagacacc	2400
gtcctgagtg tattactttt	caacagatca	aggataattt	cgctaatacg	cttgatctgc	2460
tggcgcagaa gtattccata	gagcagctga	ccacttactg	gctgcagcca	ggggatgatt	2520
ttgaggaggc tattagggta	tatgcaagg	tggcacttag	gccagattgc	aagtacaaga	2580
tcagcaaact tgtaataatc	aggaatttt	gctacattt	tggaaacggg	gccgagggtgg	2640
agatagatac ggaggtatgg	gtggcctta	gtatgtat	gataaaatg	tggccggggg	2700
tgcttggcat ggacgggggt	gttattatga	atgtaaagg	tactggcccc	aattttagcg	2760
gtacggttt cctggccat	accaacctt	tcctacacgg	tgtaaagctt	tatgggttt	2820
acaatacccg tggaaagcc	tggaccatg	taagggttcg	gggctgtgcc	tttactgtct	2880
gctggaaagg ggtgggtgt	cgccccaaa	gcagggctt	aattaagaaa	tgcctcttt	2940
aaagggtgtac ctgggtatc	ctgtctgagg	gtaactccag	ggtgcgcac	aatgtggcct	3000
ccgactgtgg ttgttccatg	ctagtaaaa	gcgtggctgt	gattaagcat	aacatggat	3060
gtggcaactg cgaggacagg	gcctctcaga	tgctgacctg	ctcgacggc	aactgttcacc	3120
tgctgaagac cattcacgt	gccagccact	ctcgcaaggc	ctggccagtg	tttgagcata	3180
acatactgac cccgtgttcc	ttgcatttg	gtaacaggag	gggggtgttc	ctaccttacc	3240
aatgcattt gactcacact	aagatattgc	ttgagccga	gagcatgtcc	aaggtaaacc	3300
tgaacgggtt gtttgcacatg	accatgaaga	tctggaggt	gtcgaggatc	gtatggagaccc	3360
gcaccagggt cagaccctgc	gagtgtggcg	gttaaacat	taggaaccag	cctgtatgc	3420
tggatgtgac cgaggagctg	aggcccgtatc	actttgggtct	ggcctgcacc	cgcgctgagt	3480
ttggctctag cgatgaagat	acagatttag	gtactgaaat	gtgtgggggt	ggcttaaggg	3540
tggaaaagaa tatataaggt	gggggtctt	tgtatgttt	tatctgttt	gcagcagccg	3600
ccggcccat gaggaccaac	tcgtttgtat	gaagcattgt	gagctcatat	ttgacaacgc	3660
gcatgcccccc atgggggggg	gtgcgtcaga	atgtatggg	ctccageatt	gatggtcgc	3720
ccgttctgcc cgccaaactct	actaccttga	cctacagac	cgtgtctgga	acgcccgttgg	3780
agactgcagc ctccggccccc	gcttcagccg	ctgcagccac	cgcccgccgg	attgtgactg	3840
actttgtttt cctgagcccg	cttgcaagca	gtgcagctt	ccgttcatcc	gcccgcgtat	3900
acaagggtgac ggctctttt	gcacaatttg	atttttgcac	ccgggaactt	aatgtcgttt	3960
ctcagcagct gtttgcacat	cgccagcagg	tttctggccct	gaaggcttcc	tccccccttca	4020
atgcgggttta aaacataat	aaaaaaccag	actctgtttt	gatttggatc	aagcaagtgt	4080
cttgcgttct ttatgttaggg	gttttgcgc	cgcggttagc	ccgggaccag	cggctcggt	4140
cgttgggggtt cctgtgtatt	ttttccagga	cggtttaaag	gtgactctgg	atgttcatag	4200
acatgggcat aagccgtct	ctgggggttga	gttagcacca	ctgcagagct	tcatgtcg	4260
gggtgggtttt gtagatgtatc	cagtgcgtac	aggagcgttgc	ggcgtgggtc	ctaaaaatgt	4320
ctttcagtag caagctgatt	gccaggggca	ggcccttggt	gtaagtgttt	acaaagcggt	4380
taagctggga tgggtgcata	cgtggggata	tgagatgtat	cttggactgt	attttttaggt	4440
tggctatgtt cccagccata	tccctccggg	gattcatgtt	gtgcagaaacc	accagcacag	4500
tgtatccgggt gcacttggga	aatttgcata	gtagctttaga	agggaaatgc	tggaaagaact	4560
tggagacgccc ttgtgacat	ccaagattt	ccatgcattt	gtccataatg	atggcaatgg	4620
gccccacggggc ggccggctgg	gcgaagatat	ttctggatc	actaacgtca	tagttgttt	4680
ccaggatgag atcgtcatag	gccattttta	caaagcgccg	cgggagggtg	ccagactcg	4740
gtataatgtt tccatccggc	ccaggggggt	agtatccctc	acagatttgc	atttcccacg	4800
ctttgcgttcc agatgggggg	atcatgtct	ccttgcggggc	atgtaaagaaa	acggtttccg	4860
gggttagggga gatcgttgg	gaagaaaacga	ggttcttgag	cagctgcac	ttaccgcagc	4920
cggtggggccc gtaaatcaca	cctattaccg	ggtcactat	gtatgttata	gagctgcagc	4980
tgccgtcatac cctgagcagg	ggggccactt	cgttaagcat	gtccctgact	cgcatgttt	5040
ccctgaccaa atccggccaga	aggcgctcgc	cgcccagcga	tagcagtct	tgcaagggaa	5100

caaagttttt	caacggtttg	agaccgtccg	ccgtaggcat	gctttgagc	gtttgaccaa	5160
gcagtccag	gcccccccac	agctcggtca	cctgctctac	ggcatctcg	tccagcatat	5220
ctccctcgtt	cgcgggttgg	ggcggcttcc	gctgtacggc	agtagtcgg	gctcgccag	5280
acggggcagg	gtcatgtctt	tccacggcg	cagggtctc	gtcagcgtag	tctgggtcac	5340
ggtaagggg	tgcgtccgg	gctgcgcgt	ggccagggtg	cgctttaggc	tggtccctgct	5400
ggtgcgtaa	cgtgcgggt	cttcgcctg	cgcgtccggc	agtaggcatt	tgaccatgg	5460
gtcatagtcc	agccccctccg	cggcggtggc	cttggcgccg	acgttgcct	tggaggaggg	5520
gcccacgag	gggcagtgc	gacttttag	ggcgtagagc	ttggcgcga	gaaataccga	5580
ttccggggag	taggcattccg	cgccgcaggc	ccgcagacg	gtctcgatt	ccacgagcca	5640
ggtagctct	ggccgttcgg	ggtaaaaaac	caggttccc	ccatgcctt	tgatgcgtt	5700
cttacctctg	gtttccatga	gccgggttcc	acgtcgggt	acgaaaaaggc	tgtccgtgtc	5760
cccgatatac	gacttgagag	gcctgttctc	gaggcgggtt	ccgcggctt	cctcgatata	5820
aaactcggac	cactctgaga	caaaggctcg	cgtccaggcc	agcacgaagg	aggctaagt	5880
ggaggggtag	cggtcgttgg	ccactagggg	gtcactcgc	tccaggggtt	gaagacacat	5940
gtgcgcctct	tccgcattca	ggaagggtat	tgggtttag	gtgtagggca	cgtgaccggg	6000
tgttctgaa	ggggggctat	aaaagggggt	ggggcgcgt	tcttcctc	tcttccgc	6060
atcgctgtct	ggcaggggcca	gctgtttttt	ttagtactcc	ctctaaaaag	cgggcgtac	6120
ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	ttgatattca	cctggcccgc	6180
ggtgatgcct	ttgagggtgg	ccgcattccat	ctgtcagaa	aagacaattt	tttttgtgtc	6240
aagctgggt	gcaaaacgacc	cgtagagggc	gttggacage	aacttggcga	tggagcgcag	6300
ggtttggttt	ttgtcgcgat	cggcgcgtc	cttggccgcg	atgttagt	gcacgtattc	6360
gcgcgcaacg	caccgcatt	cgggaaagac	ggtggcgc	tcgtcggca	ccagggtc	6420
gcgcacccg	cgttgtgc	gggtgacaag	gtcaacgcgt	gtggctaccc	ctccgcgt	6480
gcgcgtgtt	gtccgcgcga	ggcggccgc	cttgcgcgc	cagaatggcg	gtaggggtc	6540
tagctgcgtc	tctggggggg	ggctctgc	cacggtaaaag	accccgccca	gcaggcgc	6600
gtcgaagtag	tctatcttgc	atcttgc	gtctagcgc	tgcgtccat	cgccggccgc	6660
aagcgcgcgc	tcgtatgggt	ttagtgggg	acccatggc	atgggttgg	tgagcgcgga	6720
ggcgtacatg	ccgcaaatgt	cgtaaacgt	gagggctct	ctgagtttcc	caagatatgt	6780
aggtagcat	cttccaccgc	ggatgcttgc	gcccacgtaa	tcgtatagg	cgtgcgagg	6840
agcgaggagg	tccggaccga	ggttgctac	ggccggctgc	tctgctcgg	agactatct	6900
cctgaagatg	gcatgtgagt	tggatgat	ggtggacgc	tggaaagacgt	tgaagctggc	6960
gtctgtgaga	cctaccgcgt	cacgcacgaa	ggaggcgt	gagtcgc	gcttggat	7020
cagctcgccg	gtgacctgc	cgtctagg	gcagtagtcc	agggttcc	tgtatgt	7080
atactatatc	tccctttttt	ttttccacag	ctcgcgtt	aggacaaact	cttcgcgt	7140
tttccatgc	tcttggatcg	gaaaccgc	ggcctccgaa	cggtaaagac	ctagcatgt	7200
gaactgggt	acggcctgtt	aggcgc	tcccttttct	acgggttgc	cgtatgc	7260
cgccgccttc	cggagcgcagg	tgtgggtg	cgcaaaagg	tccctgacca	tgactttgag	7320
gtactggat	ttaaagtgc	tgtcgtc	tccgcctgc	tcccagagca	aaaagtccgt	7380
gcgtttttt	gaacgcggat	ttggcagg	gaaggtgaca	tcgttgaaga	gtatcttcc	7440
cgcgcgaggc	ataaaagtgc	gtgtgatgc	gaagggtccc	ggcacctcg	aacgggtt	7500
aattacctgg	gccccggc	cgatctcg	aaagccgtt	atgttgc	ccacaatgt	7560
aagtttcaag	aagcgcggg	tgccttgc	ggaaggcaat	tttttaagtt	cctcgtag	7620
gagcttca	ggggagctga	gcccgtg	tgaaagg	cagtctc	gatgagg	7680
ggaagcgac	aatgagctc	acaggtc	ggccattag	atttgc	ggtcgcgaaa	7740
ggtcctaaac	tggcgcac	tggcattt	ttctgggt	atgc	aggtaa	7800
gtcttgc	cagcggtccc	atccaagg	cgcggctt	tctcgc	cagtca	7860
aggctcatct	ccgcgcact	tcatgac	catgaagg	acgagct	tcccaaagg	7920
ccccatccaa	gtatagg	ctacatc	ggtgcaca	agacgc	tgcgagg	7980
cgagccgatc	gggaagaact	ggatctcc	ccaccaatt	gaggagtgc	tattgt	8040
gtgaaagtag	aagtccctgc	gacggccg	acactcgt	tggctttt	aaaaacgt	8100
gcagtactgg	cagcgtgc	cggcgt	atccgcac	aggttgc	gacgacc	8160
cacaaggaag	cagagtgg	atttgc	ctcgc	gggttgg	gtgttctt	8220
tacttcgg	gtttgtc	gaccgtc	ctgc	ggagttac	tggatcgac	8280
caccacccg	cgcgagccca	aagtcc	gtccgc	ggcggc	gcttgc	8340
aacatcg	agatggagc	tgtccat	ctggag	tcgcgcgt	ggtcagg	8400

gagctcctgc	aggtttacct	cgcatagacg	ggtcagggcg	cgggctagat	ccaggtgata	8460
cctaattcc	aggggctgg	tgggtggccg	gtcgatggct	tgcaagaggc	cgcatccccg	8520
cgcgcgact	acggtaccgc	gcccggcg	gtggggcccg	gggggttct	tggatgatgc	8580
atctaaaagc	ggtgacgcgg	gcgagccccc	ggaggttaggg	ggggctccgg	acccgcccgg	8640
agagggggca	ggggcacgtc	ggcgcggcgc	gcgggcagga	gctgggtctg	cgcgcgtagg	8700
ttgctggcga	acgcgacgac	gcccgggtt	atctcctgaa	tctggcgcct	ctgcgtgaag	8760
acgacggcc	cggtgagctt	gagcctgaaa	gagagttcga	cagaatcaat	ttcgggtgtcg	8820
ttgacggcgg	cctggcgcaa	aatctcctgc	acgtctcctg	atgtgtctt	ataggcgtac	8880
tcggccatga	actgctcgat	cttttcctcc	tggagatctc	cgcgtccggc	tcgcgtccacg	8940
gtggcggcga	ggtcgttgg	aatcgggccc	atgagtcgcg	agaaggcgtt	gaggcctccc	9000
tcgttccaga	cgcggctgt	gaccacccc	ccttcggcat	cgcggccgc	catgaccacc	9060
tgcgcgagat	ttagctccac	gtggggcg	aagacggcg	atgttcgcg	gcgtgaaag	9120
aggtagtga	gggtgggtgg	ggtgttctt	gccaacgaaga	agtacataac	ccagcgtcgc	9180
aacgtgatt	cgttgatatac	ccccaaaggc	tcaaggcgt	ccatggcctc	gtagaagtcc	9240
acggcgaagt	tggaaaaactg	ggagttgcgc	gcccacacgg	ttaactcctc	ctccagaaga	9300
cggatgagct	cggcgacagt	gtcgccac	tcgcgtctaa	aggctacagg	ggcctcttct	9360
tcttcttcaa	tctcctcttc	cataaggcc	tcccttctt	cttcttctgg	cgccgggtgg	9420
ggagggggga	cacggcgccg	acgacggcgc	accgggaggc	gttcgacaaa	gcgtcgatc	9480
atctcccccgc	ggcgacggcg	catggtctcg	gtacggcg	ggccgttctc	gccccggcgc	9540
agttgaaga	cggcccccgt	catgtcccg	ttatgggtt	gccccgggt	gcatgcgc	9600
agggatacgg	cgtctaaccgt	gcatctcaac	aattgttgc	taggtactcc	gcccggagg	9660
gacctgagcg	agtccgcata	gaccggatcg	aaaaacctct	cgagaaaaggc	gtetaaccag	9720
tcacagtgc	aaggtaggt	gagcacccgt	ggccggcgca	gcggccggcg	gtcggggtt	9780
tttctggcg	agggtgtct	gatgtat	ttaaagttag	cggctttag	acggcgatg	9840
gtcgacagaa	gcaccatgtc	cttgggtcg	gcctgtctaa	tgccgcaggcg	gtcgccatg	9900
ccccagctt	cgttttgaca	tcggcgagg	tcttgcgt	agtcttgcat	gaggctttct	9960
accggcaactt	cttcttcttc	ttccttctt	cctgcatactc	ttgcatactat	cgctggcg	10020
gcggcgagat	ttggccgtag	gtggccctt	tttctccca	tcgtgtgac	cccgaaagccc	10080
ctcatcggt	gaacgcggc	taggtcg	acaacgcgt	cggctaataat	ggctgtctgc	10140
acctgcgtga	gggttagactg	gaagtcatcc	atgtccacaa	agcggtgta	tgcccccgt	10200
ttgatgggt	aagtgcagtt	ggccataacg	gaccagttaa	cggcttgg	acccggctgc	10260
gagagctcg	tgtacctgag	acgcgagtaa	gccctcgagt	caaatacgt	gtcggtgca	10320
gtccgcacca	ggtactggta	tcccacaaa	aagtgcggcg	gcggctggcg	gtagagggc	10380
cagcgtaggg	tggccggggc	tccggggcg	agatcttca	acataaggcg	atgatatccg	10440
tagatgtacc	tggacatcca	ggtgatggc	ggccgggtgg	tggaggcg	cgaaaaatcg	10500
cggacgcgg	tccagatgtt	gcccggcgc	aaaaaagtct	ccatggtccg	gacgcctctgg	10560
ccggteaggc	gcccgcata	gttgcac	tagaccgtc	aaaaggagag	cctgtaaagcg	10620
ggcaacttcc	cgtggctctgg	tggataaaatt	cgcaagggt	tcatggcga	cgaccgggt	10680
tcgagccccg	tatccggccg	tccggcgta	tccatgcgt	taccggccgc	gtgtcgaacc	10740
caggtgtcg	acgtcagaca	acggggagg	gcttccttgc	gttccttcc	aggcgccgg	10800
gctgctgc	tagcttttt	ggccacttgc	cgccgcgc	gtaagcggtt	aggctggaaa	10860
gcgaaagcat	taagtggctc	gtccctgt	gcccggagggt	tatccca	gggttggatc	10920
gcgggacccc	cgttgcgat	ctcgaccgg	ccggactgc	gcgaacgggg	gtttgcctcc	10980
ccgtcatgca	agaccccgct	tgcaaattcc	tccggaaaca	gggacgagcc	ccttttttgc	11040
tttccca	tgcatccgg	gctgcggcag	atgcggcccc	ctccctcagca	cgcccaagag	11100
caagagcagc	ggcagacatg	cagggcaccc	tccctcttc	ctaccgcgtc	aggagggccg	11160
acatcccg	ttgacgcggc	agcagatgtt	gattacgaac	ccccggggcg	ccgggccccgg	11220
caactactgg	acttggagga	gggcggggc	ctggcgccgg	taggacgc	ctctcctgt	11280
cggtaaaaa	gggtgcgat	gaagcgtat	acgcgtgagg	cgtacgtcc	gcggcagaac	11340
ctgtttcg	accgcgagg	agaggagcc	gaggagatgc	gggatcgaaa	gttccacgc	11400
gggcgcg	tgcggcatgg	cctgaatgc	gagcgggttgc	tgcgcgagga	ggactttgag	11460
cccgacgc	gaaccgggat	tagtcccg	cgccgcacacg	tggccggcc	cgacactggta	11520
accgcatacg	agcagacggt	gaaccaggag	attaacttc	aaaaaagctt	taacaaccac	11580
gtgcgtacgc	ttgtggcg	cgaggaggt	gtataggac	tgatgcata	gtgggactt	11640
gtaagcgcgc	tggagcaaaa	cccaaata	aggccgtca	tggcgcagct	gttccttata	11700

cattcccgca	ctgttggatg	tggacgccta	ccagggcagc	ttgaaaatgc	acaccgaaca	15060
gggcgggggt	ggcgcaggcg	gcagcaacag	cagtggcagc	ggcgcggaag	agaactccaa	15120
cgcggcagcc	gcggcaatgc	agccgggtga	ggacatgaac	gatcatgcca	ttcgccgcga	15180
caccttgc	acacgggctg	aggagaagcg	cgctgaggcc	gaagcagcgg	ccgaagctgc	15240
cgcccccgt	gcgcaacccg	aggtcgagaa	gcctcagaag	aaaccggtga	tcaaaccct	15300
gacagaggac	agcaagaaaac	gcagttacaa	cctaataagc	aatgacagca	ccttcaccca	15360
gtaccgcagc	tggtaccttg	catacaacta	cggcgaccct	cagaccgaa	tccgctcatg	15420
gaccctgctt	tgcaactcctg	acgtaacctg	cggctcgagg	caggtctact	ggtcgttgcc	15480
agacatgatg	caagaccccg	tgaccttccg	ctccacgc	cagatcagca	acttccgg	15540
ggtgggcgccc	gagctgttgc	ccgtgcactc	caagagctc	tacaacgacc	aggccgtcta	15600
ctcccaactc	atccgcccagt	ttacctctct	gaccacgtg	ttcaatcgct	ttcccagaaa	15660
ccagatttg	gcgcggccgc	cagccccac	catcaccacc	gtcagtggaa	acgttctgc	15720
tctcacagat	cacccggacgc	taccgctgcg	caacagcatc	ggaggagtcc	agcgagtgc	15780
cattactgac	gccagacgcc	gcacctgccc	ctacgtttac	aaggccctgg	gcatagtctc	15840
gcccgcgtc	ctatcgagcc	gcacttttg	agcaagcatg	tccatcccta	tatgcggccag	15900
caataacaca	ggctggggcc	tgcgcttccc	aagaagatg	tttggcgggg	ccaagaagcg	15960
ctccgaccaa	cacccagtgc	gcgtgcgcgg	gcactaccgc	gcccctgg	gcgcgcacaa	16020
acgcggccgc	actgggcgca	ccacggtcga	tgacgcattc	gacgcgggtgg	tggaggaggc	16080
gcfgaactac	acgcccacgc	cgccaccagt	gtccacagtg	gacgcggcca	ttcagaccgt	16140
ggtgcgcgga	gccccggcgt	atgctaaaat	gaagagacgg	cgaggaggcgc	tagacgtcg	16200
ccacccggcgc	cgacccggca	ctgcccggca	acgcgcggcg	gcccctgc	ttaaccgcgc	16260
acgtcgaccc	ggccgacggg	cgcccatgcg	ggccgcgtcg	aggctggccg	cggttattgt	16320
cactgtgccc	cccagggtcca	ggcgcacgac	ggccgcggca	gcagccgcgg	ccattagtgc	16380
tatgactcag	ggtgcgcagg	gcaacgtgt	ttgggtgcgc	gactcggtt	gcgcctgcg	16440
cgtggccgt	cgacccggcc	ccccgcgaa	ctagattgc	agaaaaaact	acttagactc	16500
gtactgtt	atgtatccag	cgggggcggc	gcgcacacgaa	gtatgttca	agcgcacaaat	16560
caaagaagag	atgttccagg	tcatcgcc	ggagatctat	ggcccccgg	agaaggaaaga	16620
gcaggattac	aagcccccga	agctaaacgc	ggtcaaaaag	aaaaagaaag	atgtatgtga	16680
tgaacttgac	gacgagggtgg	aactgctgca	cgctaccgc	cccaggcgcac	gggtacagtg	16740
gaaaggtcga	cgcgtaaaaac	gtgtttgcg	acccggcacc	accgtatct	ttacgcggcgg	16800
tgagcgttcc	acccgcacct	acaagcgcgt	gtatgtatg	gtgtacggcg	acgaggacct	16860
gctttagcag	gccaacgcgc	gcctcgggg	gtttgcctac	ggaaagcggc	ataaggacat	16920
gctggcggt	cgctggacgc	agggcaaccc	aacacctagc	ctaaagcccg	taaactgc	16980
gcaggtgt	ccccgcgtt	cacccgtccg	agaaaagcgc	ggcctaataacgc	gcgagtctgg	17040
tgacttgca	ccccaccgtc	agctgtatgtt	acccaacgc	cagcgactgg	aaatgttctt	17100
ggaaaaaaatg	acccgtggaaac	ctgggcttgg	gcccggggc	cgcgactggc	caatcaagca	17160
ggtggccgc	ggactgggc	tgcagaccgt	ggacgtttag	atacccacta	ccagtagcac	17220
cagtattggc	acccgcacag	agggcatgg	gacacaacgc	tcccccgtt	cctcagcggt	17280
ggcggatgccc	gggggtgcagg	cggtcgctgc	ggccgcgtcc	aagacctcta	cgaggggtgc	17340
aacggacccg	tggatgtttc	cggtttc	cgccggcgc	ccgcgcgggtt	cgagggaaata	17400
cggccgcgc	agcgcgtac	tgcctcgaaata	tgcctatcat	ccttcatttgc	cgccctacccc	17460
cggctatcg	ggctacaccc	acccgcacag	aagacgagca	actacccgcac	gcccgaaccac	17520
cactggaaacc	cggccgcgc	gtgcgcgtcg	ccagcccggt	ctggcccccga	tttccgtgc	17580
cagggtggct	cgcgaaggag	gcaggaccct	ggtgcgtcc	acagcgcgt	accaccccg	17640
catcgttaa	aagccgttct	ttgtgggttct	tgcagatatg	gccttcacct	gcccgcctccg	17700
tttcccggt	ccgggatttcc	gaggaagaat	gcaccgttag	aggggcattgg	ccggccacgg	17760
cctgacgggc	ggcatgcgtc	gtgcgcacca	ccggccggcg	cgcgctgc	accgtcgcat	17820
gcfgccgggt	atccgtcccc	tccttattcc	actgatgc	gcccgcatttgc	gcccgcgtcc	17880
cggaaatgc	tccgtggcc	tgcaggcgca	gagacactga	ttaaaaacaa	gttgcattgt	17940
gaaaaatcaa	aataaaaaatg	ctggacttcc	acgctcgctt	ggtcctgtt	ctatgttgc	18000
gaatggaaaga	catcaacttt	gcgtctctgg	ccccgcgaca	cggtcgccgc	ccgttcatgg	18060
gaaaactggca	agatatccgc	accagcaata	tgagggtgg	cgcccttc	tggggctcgc	18120
tgtggagcgg	cattaaaaat	ttcggttcca	ccgttaagaa	ctatggc	agggccttgc	18180
acagcagcac	aggccagatg	ctgagggtata	agttgaaaga	gaaaatttc	caacaaaagg	18240
ttgttagatgg	cctggcctct	ggcattagcg	gggtgggttga	cctggccaac	caggcagtgc	18300

aaaataagat	taacagtaag	cttgatcccc	gccctcccg	agaggagcc	ccaccggccg	18360
tggagacagt	gtctccagag	gggcgtggcg	aaaagcgtcc	gcgcggcc	aggaaagaaa	18420
ctctgtgac	gcaaataagac	gagcctccct	cgta	ggca	caaggcctgc	18480
ccaccacccg	tcccacatcg	cccatggcta	ccggagtgct	gggc	cagcac	18540
cgctggac	ctgc	ccgcacaccc	agcaga	tgt	gtccgc	18600
ccgttgg	ttgt	agccgcgt	ccctgc	ccgc	ggtcg	18660
cg	gg	ggcaactggc	aaagcacact	gaac	acat	18720
gggt	ca	cgacatgt	tctgaa	taac	tg	18780
atgtatgcgt	ccatgtcg	gccagaggag	ctgt	gag	ccgcgc	18840
gatggctacc	ccttcgat	tgccgc	ctgttacat	cacat	ctcg	18900
ctcggagtac	ctgag	ggctgg	gttgc	gca	ccaggac	18960
cct	gt	ggcc	ctac	cac	gt	19020
gtcccac	gtt	ggcc	gttgc	gac	ccac	19080
caaggc	ttc	gttgc	ttgt	gat	act	19140
ctt	ttc	ttgt	taacc	ttt	ctt	19200
ctacaac	ctgg	ttgt	ccct	ttt	ttt	19260
tgct	ttg	aaat	ttt	ttt	ttt	19320
cacaat	at	ttt	ttt	ttt	ttt	19380
agct	ttt	ttt	ttt	ttt	ttt	19440
aaaggagg	ttt	ttt	ttt	ttt	ttt	19500
tcaac	ttt	ttt	ttt	ttt	ttt	19560
ttt	ttt	ttt	ttt	ttt	ttt	19620
tca	ttt	ttt	ttt	ttt	ttt	19680
ttt	ttt	ttt	ttt	ttt	ttt	19740
ttt	ttt	ttt	ttt	ttt	ttt	19800
ttt	ttt	ttt	ttt	ttt	ttt	19860
ttt	ttt	ttt	ttt	ttt	ttt	19920
ttt	ttt	ttt	ttt	ttt	ttt	19980
ttt	ttt	ttt	ttt	ttt	ttt	20040
ttt	ttt	ttt	ttt	ttt	ttt	20100
ttt	ttt	ttt	ttt	ttt	ttt	20160
ttt	ttt	ttt	ttt	ttt	ttt	20220
ttt	ttt	ttt	ttt	ttt	ttt	20280
ttt	ttt	ttt	ttt	ttt	ttt	20340
ttt	ttt	ttt	ttt	ttt	ttt	20400
ttt	ttt	ttt	ttt	ttt	ttt	20460
ttt	ttt	ttt	ttt	ttt	ttt	20520
ttt	ttt	ttt	ttt	ttt	ttt	20580
ttt	ttt	ttt	ttt	ttt	ttt	20640
ttt	ttt	ttt	ttt	ttt	ttt	20700
ttt	ttt	ttt	ttt	ttt	ttt	20760
ttt	ttt	ttt	ttt	ttt	ttt	20820
ttt	ttt	ttt	ttt	ttt	ttt	20880
ttt	ttt	ttt	ttt	ttt	ttt	20940
ttt	ttt	ttt	ttt	ttt	ttt	21000
ttt	ttt	ttt	ttt	ttt	ttt	21060
ttt	ttt	ttt	ttt	ttt	ttt	21120
ttt	ttt	ttt	ttt	ttt	ttt	21180
ttt	ttt	ttt	ttt	ttt	ttt	21240
ttt	ttt	ttt	ttt	ttt	ttt	21300
ttt	ttt	ttt	ttt	ttt	ttt	21360
ttt	ttt	ttt	ttt	ttt	ttt	21420
ttt	ttt	ttt	ttt	ttt	ttt	21480
ttt	ttt	ttt	ttt	ttt	ttt	21540
ttt	ttt	ttt	ttt	ttt	ttt	21600

tgaagtcttt	gacgtggtcc	gtgtgcacccg	gccgcaccgc	ggcgtcatcg	aaaccgtgta	21660
cctgcgaacg	cccttctcg	ccggcaacgc	cacaacataa	agaagcaagc	aacatcaaca	21720
acagctgccg	ccatgggctc	cagtgaggcg	gaactgaaag	ccattgtcaa	agatcttgg	21780
tgtgggccat	atttttggg	cacctatgac	aagcgcttc	caggcttgt	ttctccacac	21840
aagctcgct	gcgccatagt	caatacgcc	ggtcgagaga	ctggggcgt	acactggatg	21900
gcctttgcct	ggaaccgc	ctcaaaaaca	tgctacctct	tgagccctt	tggctttct	21960
gaccagcgac	tcaagcaggt	ttaccagtt	gagtagcagt	caactctcg	ccgtagcgcc	22020
attgcttctt	ccccgaccg	ctgtataacg	ctgaaaaagt	ccacccaaag	cgtacagggg	22080
cccaactcgg	ccgcctgtgg	actattctgc	tgcatgttc	tccacgcctt	tgccaactgg	22140
ccccaaactc	ccatggatca	caaaaaacc	atgaaccta	ttacccgggt	accaactcc	22200
atgctcaaca	gtccccaggt	acagccacc	ctgctcg	accaggaaca	gctctacagc	22260
ttectggagc	gcaactcgcc	ctacttcgc	agccacagt	cgcagattag	gagcgcact	22320
tctttttgtc	acttgaaaaa	catgtaaaaa	taatgtacta	gagacactt	caataaaggc	22380
aatgctttt	atttgacac	tctgggtga	ttatattaccc	ccaccccttc	cgtctcgcc	22440
gtttaaaaat	caaaggggtt	ctgcccgc	tcgctatgc	ccactggcg	ggacacgttg	22500
cgatactgg	gtttagtgc	ccacttaaac	tcagggcaca	ccatccgg	cagctcgg	22560
aagttttcac	tccacaggct	gcgcaccatc	accaacgcgt	ttagcagtc	gggcgcgat	22620
atcttgaagt	cgcagttggg	gectccgccc	tgcgcg	agttgcata	cacagggttg	22680
cagcaactgga	acactatcag	cgccgggtgg	tgca	ccagcacgt	cttgcggag	22740
atcagatccg	cgtccagg	ctccgcgtt	ctcagg	acggagtca	cttggtagc	22800
tgccttcca	aaaagggcgc	gtgcccaggc	ttttagtgc	actgcaccc	tagggcata	22860
aaaaggtgac	cgtgcccgg	ctgggcgtt	ggatacagcg	cctgcataaa	agccttgate	22920
tgcttaaaag	ccacctgagc	cttgcgc	tcagagaaga	acatccc	agacttgc	22980
gaaaactgt	tggccggaca	ggccgcgt	tgca	accttgc	gtgttggag	23040
atctgcacca	catttcggcc	ccaccgg	ttca	tggcttgc	agactgtcc	23100
ttcagcgc	gctgcccgtt	ttcgc	gcata	caatcacgt	ctccttattt	23160
atcataatgc	ttccgtgtag	acacttac	acatccattt	tctcagc	gcgggtc	23220
cacaacgc	agcccg	ctcg	tcg	ccacgtgt	cttgc	23280
tacgcctgca	ggaatgc	catcatcg	aca	cctctc	ca	23340
tgcaacccgc	ggtgc	gttc	agc	cgatcgg	cc	23400
ttgtcaggca	gtagttgaa	gttc	cc	actcagc	gg	23460
agcgcgc	cagc	gttc	gt	cgatcgg	cc	23520
ttcatcacc	taatttact	gttc	gc	ccatcg	cc	23580
ataccacgc	ccactgg	gtt	cc	tgc	cc	23640
ccatgc	tttgc	tat	cc	ctgt	cc	23700
ttttttct	cgtgttcc	ttt	tt	gttgc	cc	23760
gggcgtt	ttttttct	ttt	tt	ccgcgt	cc	23820
gggcgtt	ttttttct	ttt	tt	ccgcgt	cc	23880
atacgc	ccatcg	ttt	tt	ccgcgt	cc	23940
gacgacacgt	cctccatgt	ttt	tt	ccgcgt	cc	24000
gtttcg	gttcttct	ttt	tt	ccgcgt	cc	24060
atggagtc	tcgagaagaa	ttt	tt	ccgcgt	cc	24120
ttcac	ccgcaacgc	ttt	tt	ccgcgt	cc	24180
gaggaagt	ttatcgag	ttt	tt	ccgcgt	cc	24240
gtaccaac	aggataaaaa	ttt	tt	ccgcgt	cc	24300
gggcgggg	acgaaagg	ttt	tt	ccgcgt	cc	24360
catctgc	gc	ttt	tt	ccgcgt	cc	24420
ctcgccat	cgatgtc	ttt	tt	ccgcgt	cc	24480
cccaaaacgc	aaagaaacgg	ttt	tt	ccgcgt	cc	24540
tttgcgt	cagagg	ttt	tt	ccgcgt	cc	24600
ctatctgc	gtgca	ttt	tt	ccgcgt	cc	24660
gtcata	atatgc	ttt	tt	ccgcgt	cc	24720
gacgaga	gcgcgg	ttt	tt	ccgcgt	cc	24780
ggagtgtt	ttgaaactc	ttt	tt	ccgcgt	cc	24840
gaggtc	accc	ttt	tt	ccgcgt	cc	24900
gaggtc	accc	ttt	tt	ccgcgt	cc	

atgagtgagc tgatcgtgcg ccgtgcgcag cccctggaga gggatgcaaa tttgcaagaa 2
 caaacagagg agggcctacc cgcaaggc gacgagcagc tagcgcgtg gcttcaaacg 25,
 cgcgagcctg ccgacttggg ggagcgcac aaactaatga tggccgcagt gctcgttacc 2508,
 gtggagctt agtgcacgca gcggttctt gctgaccggg agatgcgcg caagcttagag 25140
 gaaacattgc actacaccc ttgcacaggc tacgtacgac aggcctgca gatctccaac 25200
 gtggagctt gcaaccttggt ctccctaccc ggaattttgc acgaaaaccc cttggggcaa 25260
 aacgtgcctt attccacgct caagggcag ggcgcgcgcg actacgtccg cgactgcgtt 25320
 tacttatttc tatgtcacac ctggcagac gccatggcg tttggcagca gtgcttggag 25380
 gagtgcaccc tcaaggagct gcagaaactg ctaaaagcaaa acttgaagga cctatggacg 25440
 gccttcaacg agcgcctccgt ggccgcgcac ctggcggaca tcattttcc cgaacgcctg 25500
 cttaaaaccc tgcacacaggg tctgcacagc ttaccagtc aaagcatgtt gcagaacttt 25560
 aggaacttta ttcttagagc ctcaggaaatc ttggccgcgc cctgtgtgc acttccctagc 25620
 gactttgtgc ccatthaatgc cgcgaatgc cctccgcgc tttggggca ctgtaccc 25680
 ctgcagctag ccaactaccc tgcctaccac tctgacatcaa tggaaagacgt gagcgggtgac 25740
 ggtctactgg agtgcactg tcgctcaac ctatgcaccc cgacccgtc cctgggttgc 25800
 aattcgcacg tgcataacgca aagtcaaatt atcggtaccc ttgagctgca gggccctcg 25860
 cctgacgaaa agtccgcggc tccgggggtt aaactcactc cgggggtgtg gacgtcggt 25920
 tacccctgcgca aatttgcaccc tgaggactac cacggccacg agatttagttt ctacgaagac 25980
 caatcccgcc cgcacaaatgc ggagcttacc gcctgcgtca ttacccaggg ccacattctt 26040
 ggccaattgc aagccatcaa caaagccgcg caaggtttc tgctacgaaa gggacggggg 26100
 gtttacttgg accccccagtc cggcgaggag ctcacccca tcccccggcc gccgacggcc 26160
 tatcagcagc agccgcgggc cttgttcc caggatggca cccaaaaaaa agctgcagct 26220
 gcccggccca cccacggacg aggagaata ctgggactc cagggcaggg aggttttgg 26280
 cgaggaggag gaggacatga tggaaagactg ggagagccta gacggagaaatc cttccgaggt 26340
 cgaagaggtg ttagacgaaa caccgtcacc ctggcgtgc tttccctcgcc cggcgcccc 26400
 gaaatcggca accgggttcca gcatggctac aacccctcgct cttcaggcgc cgccggcact 26460
 gcccgttgcg cgcacccaaacc gtagatggg caccactggg accagggccg gtaagtccaa 26520
 gcagccggcc cctgttagccc aagagcaaca acagcgcac ggctaccgct catggcgccg 26580
 gcacaagaac gccatagttt cttgttgcgca agactgtggg ggcaacatct cttcggccc 26640
 cccgttctt ctcttccatc acggcggtgc cttcccccgt aacatcctgc attactacc 26700
 tcatctctac agccctactc gtagatgggg cagggcgcgc ggcagcaaca gcagccggcca 26760
 cacagaagca aaggcgaccg gatagcaaga ctctgacaaa gcccaagaaa tccacagccg 26820
 cggcagcagc aggaggagga ggcgtgcgtc tggcccccgg cgaacccggta tcgaccggcg 26880
 agcttagaaa caggatttt cccactctgt atgttatatt tcaacagagc agggccaaag 26940
 aacaagagct gaaaataaaa aacaggcttc tgctatccct caccggcgcg tgcctgtatc 27000
 aaaaaagcga agatcagctt cggcgcacgc tggaaagacgc ggaggcttc ttcagtaat 27060
 actgcgcgtt gactcttaag gacttagttc gccccttc tcaaattttaa ggcgcggaaac 27120
 tacgtcatct ccagcggcca caccggcgc cagcacctgt cgtcagcgtt attatgagca 27180
 agggaaattcc caccggcctac atgtggagtt accagccaca aatgggactt gcccgttgc 27240
 ctgccccaga ctactaacc cgaataaaatc acatgagcgc gggacccac atgatatccc 27300
 gggtaacccgg aatccgcgc caccggaaacc gaatttctt ggaacaggcg gctattacca 27360
 ccacacccctcg taataaacctt aatccccgtt gttggcccg tggccgttgg taccaggaaa 27420
 gtcccgttcc caccactgtt gtaacttccca gagacggccca ggcggaaatc cagatgacta 27480
 actcaggggc gcaacttgcg ggcggcttgc gtacacagggt ggcgtgcggc gggcaggta 27540
 taactccaccc gacaatcaga gggcgaggta ttcaacttgc gacgcgttgc tgtagcttcc 27600
 cgcttggctt ccgtccggac gggacatttc agatcggcgg cgcggccgtt ctttattca 27660
 cgcctcgta gcaatccata actctgcaga cttgttccctc tgacccgcgc tctggaggca 27720
 ttggaaactct gcaattttt gaggatggg tgccatcggt ctactttaac cccttctcg 27780
 gaccccccgg ccactatccg gatcaatcc ttccacttgc tgacccgcgtt aaggactcgg 27840
 cggacccgtt cgcacttgcgatggatg ttaacttgcg aggacggacgca actgcgcctg aaacacccctgg 27900
 tccacttgcg cccggccacaag tgcttgcgc ggcacttgcg tgatgttgc tactttgaat 27960
 tgcccgaggaa tcaatccatc gggccggcgc acggcgttccg gcttaccgccc cagggagac 28020
 ttggccgttgc cctgttgcg gatgttaccc agccggccctt gcttagttgg cgggacgggg 28080
 gaccctgtgt ttcacttgcg atttgcact gtcctaaatc ttgattacat caagatctt 28140
 gttgcctatct ctgtgttgcgataataatc acagaaatccaaatataactg gggcttcttat 28200

cgccatcctg	taaacgccac	cgtcttacc	cgcacaagca	aaccaaggcg	aaccccttac	28260
ggtactttt	acatctctcc	ctctgtgatt	tacaacagtt	tcaacccaga	cggagtgtgat	28320
ctacgagaga	acctctccga	gctcagctac	tccatcagaa	aaaacaccac	cctcccttacc	28380
tgccggaaac	gtacgagtgc	gtcaccggcc	gctgcaccac	acccatccggc	tgaccgtaaa	28440
ccagactttt	tccggacaga	cctcaatac	tctgtttacc	agaacaggag	gtgagcttag	28500
aaaaccctta	gggttatttagg	ccaaaggcg	agctactgtg	ggggtttatga	aaatctcaag	28560
caactctacg	ggcttattcta	attcaggtt	ctctagaatc	gggggtgggg	ttattctctg	28620
tcttgcattt	ctcttttattc	ttatactaac	gcttctctgc	ctaaggctcg	ccgcctgctg	28680
tgtgcacatt	tgcatttattt	gtcagctttt	taaacgcgtt	gtgcgcacc	caagatgatt	28740
aggatcataa	tccttaggtt	actcaccctt	gctcagccc	acggtaccac	ccaaaagggtg	28800
gatttttaagg	agccagccctg	taatgttaca	ttcgagctg	aagctaata	gtgcaccact	28860
cttataaaat	gcaccacaga	acatgaaaag	ctgcttattc	gccacaaaaa	caaaatttggc	28920
aagtatgctg	tttatgttat	ttggcagcca	ggtgacacta	cagagtataa	tgttacagtt	28980
ttccagggtt	aaagtcataa	aactttatg	tatacttttc	cattttatga	aatgtgcgac	29040
attaccatgt	acatgagcaa	acagtataag	tttgggcccc	cacaaaattt	tgtggaaaac	29100
actggcactt	tctgtgcac	tgctatgcta	attacagtgc	tcgcttttgt	ctgtacccta	29160
ctctatattt	aatacaaaag	cagacgcagc	tttattgagg	aaaagaaaat	gccttaattt	29220
actaaggtaa	aaagctaata	tcaccactaa	ctgtttact	cgctgctgc	aaaacaaaatt	29280
caaaaagttt	gcattataat	tagaatagga	tttaaacc	ccggtcattt	cctgctcaat	29340
accattcccc	tgaacaattt	actctatgtg	ggatatgtc	cagcgctaca	accttgaagt	29400
caggcttctt	ggatgtcagc	atctgactt	ggccagcacc	tgtccgcgg	atttgttcca	29460
gtccaaactac	agcgacccac	cctaacaagag	atgaccaaca	caaccaacgc	ggccgcgcgt	29520
accggactta	catctaccac	aaatacaccc	caagttctg	cctttgtcaa	taactgggat	29580
aacttggca	tgtgggtt	ctccatagcg	cttatgttt	tatgccttat	tattatgtgg	29640
ctcatctgt	gcctaaagcg	caaacgcg	cgaccacca	tctatagtc	catattgtg	29700
ctacacccaa	acaatgtgg	aatccataga	ttggacggac	tgaaacacat	gttctttct	29760
cttacagtat	gattaaatga	gacatgattc	ctcgagttt	tatattactg	acccttggtg	29820
cgctttttt	tgcgtgtcc	acattggctg	cggttctca	catcgaagta	gactgcattc	29880
cagccttcac	agtcttattt	ctttacggat	ttgtcaccct	cacgctcatc	tgcagcctca	29940
tcactgttgt	catcgccctt	atccagtgc	ttgactgggt	ctgtgtgcgc	tttgcata	30000
tcagacacca	tcccagttac	agggacacca	ctatagctg	gtctttttaga	attttaat	30060
tatgaaattt	actgtgactt	ttctgtgtat	tatttgcacc	ctatcgctgt	tttgcatttt	30120
gacctccaag	cctaaaagac	atatactatg	cagatttact	cgatatagga	atattccaag	30180
ttgctacaat	gaaaatttggc	atcttccga	agcttggta	tatgcata	tctctgttat	30240
ggtgttctgc	agtaccatct	tagcccttagc	tataatatccc	tacattgaca	ttgctggaa	30300
acgaatagat	gccatgaacc	acccaactt	ccccgcgc	gtatgc	cactgcaaca	30360
agttgttgc	ggccggcttt	tcccagccaa	ttagcctcgc	cccacttctc	ccaccccccac	30420
tgaaaatcagc	tactttaatc	taacaggagg	agatgactga	caccctagat	ctagaaatgg	30480
acggaattat	tacagagcg	cgcctgttag	aaagacgcag	ggcagcggcc	gagaacacgc	30540
gcatgaatca	agagctccaa	gacatggta	acttgcacca	gtgaaaagg	gttatcttt	30600
gtctggtaaa	gcaggccaaa	gtcacctacg	acagtaatac	caccggacac	ccgccttagct	30660
acaagggtcc	aaccaagcgt	cagaatttgg	ttgtcatgtt	gggagaaaag	cccattacca	30720
taactcagca	ctcggttagaa	accgaaggct	gcattcactc	accttgc	ggacctgagg	30780
atctctgcac	ccttattttt	accctgtgcg	gtctcaaaaga	tcttattttt	ttaactaat	30840
aaaaaaaaat	aataaagcat	cacttacta	aaatcagtt	gcaattttct	gtccagttt	30900
ttcagcagca	cctcccttgc	ctccctccag	ctctggatt	gcagcttct	cctggctgc	30960
aactttctcc	acaatctaaa	tggaatgtca	tttgcctct	gttcctgtcc	atccgcaccc	31020
actatcttca	tgttgttgc	gatgaagcgc	gcaagaccgt	ctgaagatac	tttcaacccc	31080
gtgtatccat	atgacacacg	aaccggctt	ccaaactgtgc	cttttcttac	tttccctt	31140
gtatccccca	atgggttca	agagagtccc	cttgggtac	tcttttgcg	cctatccgaa	31200
cctctagttt	cctccaaatgg	catgcttgcg	ctaaaaatgg	gcaacggcct	ctctctggac	31260
gaggccggca	accttaccc	ccaaaatgtt	accactgtg	gcccaccc	caaaaaaaacc	31320
aagtcaaaaca	taaaccttgg	aatatctgca	cccttcacag	ttacctcaga	agccctaact	31380
gtggctgccc	ccgcaccc	aatggtcg	ggcaacacac	tcaccatgca	atcacaggcc	31440
ccgcttaacc	tgcacgactc	caaacttagc	attgcccaccc	aaggaccct	cacagtgtca	31500

actatcactg	cctcaccccc	tctaactact	gccactggta	gcttgggcat	tgacttgaaa	31560
gagcccattt	atacacaaaa	tggaaaacta	ggactaaagt	acggggctcc	tttgcatgt	31620
acagacgacc	taaacacttt	gaccgttagca	actggtccag	gtgtgactat	taataatact	31680
tccttgc当地	ctaaagttac	tggagccttg	ggtttgatt	cacaaggcaa	tatgcaactt	31740
aatgttagcc	gaggactaag	gattgattct	caaaacagac	gccttatact	tgatgttagt	31800
tatccgttt	atgctaaaa	ccaactaaat	ctaagactag	gacagggccc	tcttttata	31860
aactcagccc	acaacttgg	tattaactac	aacaaggcc	tttacttgtt	tacagcttca	31920
aaaatttcca	aaaagcttg	ggtaaactca	agcactgcca	aggggtttagt	gtttgacgct	31980
acagccatag	ccattatgc	aggagatgg	cttgaattt	gttcacccaa	tgcaccaaac	32040
acaaatcccc	tcaaaacaaa	aattggccat	ggccttagaa	tttgcattaa	caaggctatg	32100
gttccttaaac	taggaacttgg	ccttagttt	gacagcacag	gtgcattac	agtaggaaac	32160
aaaaataatg	ataagctaac	tttgggacc	acaccagtc	catctccaa	ctgtgactaa	32220
aatgcagaga	aagatgctaa	actcacttt	gtcttaacaa	aatgtggcag	tcaaaaactt	32280
gctacagttt	cagttttggc	tgtaaaggc	agtttgctc	caatatctgg	aacagttcaa	32340
agtgtctatc	ttattataag	atttgacgaa	aatggagtgc	tactaaacaa	ttccttcctg	32400
gaccgagaat	attgaaactt	tagaaatgg	gatcttactg	aaggcacagc	ctatacaaac	32460
gctgttgat	ttatgcctaa	cctatcagct	tatccaaaat	ctcacggtaa	aactgccaaa	32520
agtaacattt	tcagtcagtt	ttacttaaac	ggagacaaaa	ctaaacctgt	aacactaacc	32580
atttacactaa	acggatcac	ggaaacaggaa	gacacaactc	caagtgcata	ctctatgtca	32640
ttttcatggg	actggcttgg	ccacaactac	attaatggaa	tatttgcac	atccctttac	32700
acttttcat	acattgcccc	agaataaaga	atcgtttgc	ttatgttca	acgtgtttt	32760
ttttcaattt	cagaaaattt	caagtcat	ttcattcagt	agtagatccc	caccaccaca	32820
tagcttatac	agatcaccgt	accttaatca	aactcacaga	acccttagtat	tcaacctgccc	32880
acctccctcc	caacacacag	agtacacagt	cctttctccc	cggtggcct	taaaaagcat	32940
catatcatgg	gtaacagaca	tattcttag	tgttatattc	cacacggttt	ctgtcgagc	33000
caaacgc当地	tcagtgat	taataaactc	ccggggcagc	tcacttaagt	tcatgtcgct	33060
gtccagctgc	tgagccacag	gctgtgtcc	aacttgcgtt	tgcttaacgg	gcggcgaagg	33120
agaagtccac	gcctatctgg	gggttagatc	ataatctgtc	atcaggatag	ggcgggtgtg	33180
ctgcagcagc	gcggcaataaa	actgtgcgg	cgccgc当地	gtcctgcagg	aatacaacat	33240
ggcagtggtc	tcctcagcga	tgattcgcac	cgccccgcagc	ataaggcgcc	ttgtctccg	33300
ggcacagcag	cgcacccctga	tctcacttac	atcagcacag	taactgcagc	acagcaccac	33360
aatattgttc	aaaatcccac	agtgcagggc	gctgtatcca	aaagctcatgg	cggggaccac	33420
agaacccacg	tggccatcat	accacaagcg	caggtagatt	aaagtggcgac	ccctcataaaa	33480
cacgctgac	ataaacattt	cctcttttgg	catgttgc当地	ttcaccaccc	ccgggtacca	33540
tataaaccctc	tgatcaaaca	tgccgc当地	caccaccatc	ctaaaccagc	tggccaaaac	33600
ctgcccggc	gctataact	gcagggaaacc	gggactggaa	caatgacagt	ggagagccca	33660
ggactctgtaa	ccatggatca	tcatgtctgt	catgatatac	atgttggcac	aacacagcga	33720
cacgtgcata	cacttctca	ggatttacaag	ctcctccccc	gttagaaacca	tatcccagg	33780
aacaaccat	tcctgatca	gcgttataacc	cacactcgag	ggaagggactc	gcacgtact	33840
cacgttgtc	attgtcaaaag	tggttgc当地	gggcagcagc	ggatgtatct	ccagtatgtt	33900
agcgcgggtt	tctgtctaa	aaggaggtag	acgatccctt	ctgtacggag	tgccggagaa	33960
caaccgagat	cgtgttggc	gtagtgtcat	gccaatgg	acggccggacg	tagtcatatt	34020
tcctgaagca	aaaccagg	cgggcgtgac	aaacagatct	gctgtctccgg	tctcgccgct	34080
tagatcgctc	tgtgttag	ttgttagtata	tccactctt	caaagcatcc	aggcgcccc	34140
tggcttcggg	ttctatgtaa	actccttcat	gcccgc当地	cctgataaca	tccaccaccc	34200
cagaataa	acaccccagc	caaccttacac	atcgttctg	cgagtccac	acgggaggag	34260
cgggaagagc	tggaaagaaacc	atgttttttt	tttatttcca	aaagattatc	aaaaacctca	34320
aaatgaagat	cttataagt	aacgcgtcc	cctccgg	cggtgtcaaa	ctctacagcc	34380
aaagaacaga	taatggcatt	tgttagatc	tgccacaatgg	cttccaaaag	gcaaaacggcc	34440
ctcacgttca	agtggacgt	aaggcttacac	ccttcagggt	gaatctctc	tataaaccat	34500
ccagcacctt	caaccatgcc	caaataattt	tcatctcgcc	accttctcaa	tatctcttca	34560
agcaaaatccc	gaatattaag	tccggccatt	gtaaaaatct	gctccagagc	gcccctcc	34620
ttcagctca	agcagcgaat	catgattgc	aaaattcagg	tttctcacag	acctgtataa	34680
gattcaaaag	cggaacattt	acaaaaataac	cgcgatcccg	tagtccccc	cgcagggcca	34740

gctgaacata	atcggtcagg	tctgcacgga	ccagcgccgc	cacttcccc	ccaggaacct	34860
tgacaaaaga	acccacactg	attatgacac	gcatactcg	agctatgcta	accagcgtag	34920
ccccgatgt	agctttgttg	catggcgcc	gatataaaaat	gcaagggtct	gctaaaaaaa	34980
tcaggcaaaag	cctcgccaa	aaaagaaaagc	acatcgtagt	catgctcatg	cagataaagg	35040
caggtaaagt	ccggaaccac	cacagaaaaa	gacaccattt	ttctctcaaa	catgtctcg	35100
ggtttctgca	taaacacaaa	ataaaataac	aaaaaaaacat	ttaaacat	gaagcgtgc	35160
ttacaacagg	aaaaacaacc	cttataagca	taagacggac	tacggccat	ccggcgtgac	35220
cgtaaaaaaa	ctgtcaccg	tgattaaaaa	gcaccaccga	cagcttctcg	gtcatgtccg	35280
gagtctataat	gtaagactcg	gttttttttttt	cagggttatt	catcggtcag	tgctaaaaag	35340
cgaccgaaat	agcccggggg	aatacatacc	cgccaggcgta	gagacaacat	tacagcccc	35400
ataggaggtt	taacaaaatt	aataggagag	aaaaacacat	aaacacctga	aaaacccctcc	35460
tgcccttagca	aaatagcacc	ctcccgctcc	agaacacat	acagcgcttc	acagcggcag	35520
cctaacatgc	agcttttacca	gttttttttttt	aaaccttatta	aaaaacacc	actcgacacg	35580
gcaccaggctc	aatcgttac	agtgttttttt	aggccaaat	gcagagcgt	tatataatagg	35640
actaaaaaat	gacgttaacgg	ttaaagtcc	caaaaaacac	ccagaaaaacc	gcacgcgaac	35700
ctacgcccag	aaacgaaagc	caaaaaaccc	acaacttctt	caaatacgta	cttcgtttt	35760
cccacgttac	gtaacttccc	attttaagaa	aactacaatt	cccaacacat	acaagttact	35820
ccgccttaaa	acctacgtca	ccggcccccgt	tcccacgccc	cgcccccgt	cacaaactcc	35880
acccctcat	tatcatattt	gcttcaatcc	aaaataaggt	atattatttga	tgat	35935

<210> 9

<211> 35935

<212> DNA

<213> Adenovirus serotype 5

<400> 9

catcatcaat	aatatacctt	attttggatt	gaagccaata	tgataatgag	gggggtggagt	60
ttgtgacgtg	gcgcggggcg	tgggaacggg	gcgggtgacg	tagtagtgt	gcggaagtgt	120
gatgttcaa	gtgtggcgga	acacatgtaa	gcgacggatg	tggaaaaagt	gacgtttttg	180
gtgtgcggcg	gtgtacacag	gaagtgcacaa	ttttcgccg	gtttttaggcg	gatgttgtag	240
taaattttgg	cgttaaccgag	taagatttgg	ccattttcgc	ggggaaactg	aataagagga	300
agtggaaatct	gaataatttt	gtgttactca	tagcgcgtaa	tatttgtcta	ggggccgcggg	360
gacttttggacc	gtttacgtgg	agactcgccc	agggtttttt	ctcagggttt	ttccgcgttc	420
cgggtcaaag	ttggcggtttt	attattatag	tcaagtcgacg	tgtagtgtat	ttatacccccgg	480
tgagtttcc	aaagaggccac	tcttgagtgc	cagcgagtag	agtttttcc	tccgagccgc	540
tccgcacacc	ggactgaaaaa	tgagacat	tatctgcac	ggaggttta	ttacccgaaga	600
aatggccccc	agtttttttgg	accagctgtat	cgaagggat	ctggctgata	atcttccacc	660
tcttagccat	tttgaaccac	ctacccttca	caactgtat	gatttagacg	tgacggcccc	720
cgaagatccc	aaacgaggagg	cggtttcgca	gatttttccc	gactctgtaa	tgttggcggt	780
gcagggaaagg	attgacttac	tcactttcc	gccccggccc	ggtttccgg	agccgcctca	840
cctttcccg	cagcccgagc	agccggagca	gagaggctt	ggtccggttt	ctatgccaaa	900
ccttgtaccg	gaggtgtatcg	atcttacctg	ccacgaggct	ggcttccac	ccagtgcacg	960
cgaggatgaa	gagggtgagg	agtttggttt	agattatgt	gagcacc	ggcacgggtt	1020
caggcttctgt	cattatcacc	ggaggaatac	gggggaccac	gatattatgt	gttcgtttt	1080
ctatatgagg	acctgtggca	tgtttgtcta	cagtaagtga	aaattatggg	cagtgggtga	1140
tagagtgg	ggtttgggtt	ggttaatttt	tttttaattt	ttacagttt	gtgtttttaaa	1200
gaattttggta	ttgtgttttt	tttaaaaaggt	cctgtgtctg	aacctgagcc	tgagcccgag	1260
ccagaaccgg	agcctgcaag	acctaccgc	cgttctaaaaa	tggcgctgc	tatcttgaga	1320
cgccccgacat	cacctgtgtc	tagagaatgc	aatagtagta	cgatagctg	tgactccgt	1380
ccttctaaaca	cacccctctga	gatacaccgg	gtggcccccgc	tgtgccccat	taaaccagg	1440
gccgtgagag	ttggtggggcg	tcgcccggct	gtggaaatgt	tcgaggactt	gcttaacgag	1500
cctggccaaac	cttggactt	gagctgtaaa	cgccccagac	cataagggt	aaacctgtga	1560
ttgcgtgtgt	ggttaacgccc	tttggtttgc	gaatggatgt	atgtaaat	aataaagggt	1620
gagataatgt	ttaacttgca	tggcggttta	aatggggcg	ggcttaaagg	gtatataat	1680
cgccgtggc	taatcttgg	tacatctgac	ctcatggagg	cttggagtg	tttggaaat	1740

tttctctgtc	tgcgttaactt	gctggAACAG	agctctaaca	gtaccttgg	gttttggagg	1800
tttctgtgg	gctcatccc	ggcaaAGTT	gtctgcgaa	ttaaggagg	ttacaagtgg	1850
gaatttgaag	agcttttgg	atcctgtgg	gagcttgg	attctttgg	tctgggtcac	1920
caggcgctt	tccaaagagaa	ggtcatcaag	actttggatt	tttccacacc	ggggcgcgct	1980
gcggctgtc	ttgcctttt	gagttttata	aaggataat	ggagcgaaga	aaccatctg	2040
agcgggggg	acctgtgtt	tttctggcc	atgcatctgt	ggagagcggt	tgtgagacac	2100
aagaatccgc	tgctactgtt	gtcttccgtc	cgccggcga	taataccgac	ggaggagcag	2160
cagcagcage	aggaggaagc	caggcggcg	cgccaggcgc	agagcccatg	gaacccgaga	2220
gccccctgg	accctggga	atgaatgtt	tacaggtggc	tgaactgtat	ccagaactga	2280
gacgcattt	gacaattaca	gaggatggc	aggggtaaa	gggggttaag	aggggcggg	2340
gggcttggta	ggctacagag	gaggcttaga	atctagcttt	tagcttaatg	accagacacc	2400
gtcctgagt	tattactttt	caacagatac	aggataattt	cgctaatgag	cttgatctgc	2460
tggcgagaa	gtattccata	gagcagctg	ccacttactg	gctgcagcca	ggggatgatt	2520
ttgaggaggc	tattagggt	tatgcaaagg	ttggacttag	gccagattgc	aagtacaaga	2580
tcagcaaact	tgtaaatatac	aggaattgtt	gctacattt	tggaaacggg	gcccagggtgg	2640
agatagatac	ggaggatagg	gtggccttta	gatgttagcat	gataaatatg	tggccggggg	2700
tgcttggcat	ggacgggggt	gttattatga	atgtaaagg	tacttggcccc	aattttagcg	2760
gtacgggtt	cctggccaaat	accaaccta	tcctacacgg	tgtaaagttc	tatgggttta	2820
acaataactg	tgtggaaagcc	tggaccgat	taagggttgc	gggctgtggc	tttactgtct	2880
gctggaaagg	gggtgtgtgt	cggcccaaaa	gcagggttcc	aattaagaaa	tgccttcttg	2940
aaaggtgtac	cttgggttac	ctgtctgagg	gtaactccag	ggtgcggccac	aatggccct	3000
ccgactgtgg	ttgcttcatg	ctagtggaaa	gcgtggctgt	gattaagcat	aacatggtat	3060
gtggcaactg	cgaggacagg	gcctctcaga	tgtgacccctg	ctcggaacggc	aactgtcacc	3120
tgctgaagac	cattcacgt	gccagccact	ctcgcaaggc	ctggccagtg	tttgagcata	3180
acatactgac	ccgcgtttcc	ttgcattttg	gtaacaggag	gggggtgttc	ctaccttacc	3240
aatgcaattt	gagtcacact	aagatatttgc	ttgagccgaa	gagcatgtcc	aaggtaacc	3300
tgaacgggt	gttttgcacatg	accatgaaga	tctggaaagg	gctgagggtac	gatgagaccc	3360
gcaccagggt	cagaccctgc	gagtgtggcg	gtaaacat	taggaacccag	cctgtgtatc	3420
tggatgtac	cgaggagctg	aggcccgatc	acttgggtgc	ggcctgtcc	cgcgctgagt	3480
ttggctctag	cgatagaaat	acagatttgc	gtacttggaaat	gtgtgggggt	ggcttaaagg	3540
tggaaagaa	tatataaggt	gggggtctt	tgtatgttt	tatctgttt	gcagcagccg	3600
ccgcccacat	gagcaccaac	tcgtttgtat	gaagcatat	gagctcatat	ttgacaacgc	3660
gcatgcccc	atggccggg	gtgcgtcaga	atgtgtatgg	ctccagcatt	gatggtcgccc	3720
ccgtccctcc	cgaaactct	actacatttgc	cctacagagac	cgtgtctgg	acgcgttgg	3780
agactgcacg	ctccggccgc	gttccagccg	ctgcagccac	ccccggggg	attgtgactg	3840
actttgttt	cctgagcccg	tttgcacatc	gtgcagcttc	cogtcatcc	gcccgcgtat	3900
acaagggtac	ggcttcttttgc	gcacaatttgc	attctttgc	ccgggaactt	aatgtcgttt	3960
ctcgcacgt	gttgcacatc	cgccagcagg	tttctggccct	gaaggcttcc	tcccctccca	4020
atgcgttta	aaacataaat	aaaaaaacccat	actcttttttgc	gatttggatc	aagcaagtgt	4080
cttgcgtct	ttatattaggg	gttttgcggc	cgccgttgg	ccgggaccag	cggtctcggt	4140
cgttgggggt	cctgtgttatt	ttttccaggaa	cgtgtttaag	gtacttgcgg	atgttcagat	4200
acatggccat	aagcccgatc	ctgggggtgg	ggtagacacca	ctgcagacgt	tcatgctgcg	4260
gggtgggtt	gtagatgtatc	cagtcgttgc	aggagcgctg	ggcgttgg	ctaaaaatgt	4320
ctttcagtag	caagtcgtt	gccaggggca	ggcccttgg	gtaaagtgttt	acaaagcggt	4380
taagctgg	tgggtgcata	cggtgggata	tgagatgtatc	cttggactgt	attttttaggt	4440
tggctatgtt	cccagccata	tccctccggg	gattcatgtt	gtgcagaacc	accagcacag	4500
tgtatccgtt	gcaatggga	aatttgcata	gtagcttgc	agggaaatgcg	tggaaagaact	4560
tggagaccc	cttgcgtatc	ccaagattttgc	ccatgcatttgc	gtccataatgc	atggcaatgg	4620
gccccacgg	ggcggccctgg	gcaagatatttgc	ttctggatc	actaaacgtca	tagttgtt	4680
ccaggatgag	atcgatcat	ggcatttttgc	caaaaggcgg	ggggagggttgc	ccagactgcg	4740
gtataatgtt	tccatccggc	ccaggggcggt	aggatccctgc	acagatttgc	atttccacg	4800
ctttgagttc	agatgggggg	atcatgttgc	cctgcggggc	gatagagaaaa	acggtttccg	4860
gggttaggg	gatcagctgg	gaagaaagca	ggttccgtag	cagctgcgcac	ttaccgcac	4920
cggtggggcc	gtaaatcaca	catttacccg	ggtgcaactg	gtatgttgc	gagctgcac	4980
tgccgtcatac	cctgagcagg	ggggccacttgc	cgtaagcat	gtccctgact	cgcatgttt	5040

ccctgaccaa	atccgcccaga	aggcgctcgc	cgcggcggcga	tagcagttct	tgcaagggaaag	5100
caaagtttt	caacgggttg	agaccgtccg	cgtaggcatt	gcttttgagc	gtttgaccaa	5160
gcagttccag	gcgggtccac	agctcggtca	cctgtctac	ggcatctcg	tccagcatat	5220
ctccctcgtt	cgccgggttgg	ggcggttttc	gctgtacggc	agtagtcgg	gctcgtccag	5280
acggggccagg	gtcatgttct	tccacggcg	cagggtcctc	gtcagcgtag	tctgggtcac	5340
ggtgaagggg	tgcgtccgg	gctgcgcgt	ggcagggtg	cgcttgcggc	tggctctgt	5400
ggtgctgaag	cgctgcccgt	cttcgcctg	cgctcgcc	agtaggcatt	tgaccatgt	5460
gtcatagtc	agccctcccg	cgcggtggcc	cttggcgcgc	agttgcct	tggaggaggc	5520
gcccgcacgg	gggcagtgca	gacttttgag	ggcgtagagc	ttggcgcga	gaaataccga	5580
ttccggggag	taggcatccg	cgccgcaggc	cccgcagac	gtctcgcat	ccacgagcca	5640
ggtgagctct	ggccgttccgg	ggtaaaaaac	cagtttccc	ccatgtctt	tgatgcgtt	5700
cttacccctg	gttccatga	gcccgttcc	acgtcggt	acgaaaggc	tgtccgtgtc	5760
cccgatatac	gacttgagag	gcttgcctc	gagcggtt	cccggttct	cctcgtatag	5820
aaactcggac	cactctgaga	caaaggctcg	cggtccaggc	agcacgaagg	aggtaatgt	5880
ggaggggtag	cggtcgttgt	ccactaggg	gtccactcgc	tccagggtgt	gaagacacat	5940
gtgcgcctct	tcggcatcaa	ggaaggtat	tggttgtat	gtgttagggca	cgtgaccggg	6000
tgttcctgaa	ggggggctat	aaaaggggt	gggggcgcgt	tcgtcctcac	tctttccgc	6060
atcgctgtct	gcgagggcca	gcttggggg	tgagttactc	ctctgaaaag	cggcatgac	6120
ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	ttgatattca	cctggcccgc	6180
ggtgatgcct	ttgagggtgg	ccgcatccat	ctgttcagaa	aagacaatct	tttgggtgtc	6240
aagttggtg	gcaaacgacc	cgtagagggc	gttggacagc	aacttggca	tggagcgcag	6300
ggtttgggtt	ttgtcgcgat	cgcgccgtc	cttggccgcg	atgttttagct	gcacgtattc	6360
gcgcgcac	caccgcatt	cgggaaagac	gggtgtgcgc	tcgtcggca	ccagggtgcac	6420
gcgcaccc	cggttgcga	gggtgacaag	gtcaacgcgt	gtggctacct	ctccgcgtag	6480
gcgcgcgttgc	gtccagcaga	ggcggccgc	cttgcgcag	cagaatggcg	gtaggggtgc	6540
tagctgcgc	tcgtccgggg	ggtctcgctc	cacggtaaag	accccgggca	gcaggcgcgc	6600
gtcgaagtag	tctatcttgc	atccctgcaa	gtctagcgc	tgctgcat	cgccggccgc	6660
aagcgccgc	tcgtatgggt	tgagtgggg	acccatggc	atgggggtgg	tgagcgcgga	6720
ggcgtacatg	ccgcaaatgt	cgtaaacgta	gaggggtct	ctgagttattc	caagatatgt	6780
agggtagcat	cttccaccgc	ggatgctggc	gcgcacgtaa	tcgtatagtt	cgtgcgaggg	6840
agcgaggagg	tcgggaccga	ggttgcac	ggccggctgc	tctgctcga	agactatctg	6900
cctgaagatg	gcatgtgagt	tggatgat	ggttggacgc	tggaaagacgt	tgaagctggc	6960
gtctgtgaga	cctaccgcgt	cacgcaccaa	ggaggcgtag	gagtcgcga	gcttggtgac	7020
cacgtcgccg	gtgacactgca	cgtctaggc	gcagtagtcc	agggtttct	tgtatgtgtc	7080
atacttatac	tgcccttttt	tttccacag	ctcgccgtt	aggacaact	cttcgcggc	7140
tttccatgtac	tcttggatcg	gaaacccgtc	ggcctccgaa	cggtaagac	ctagcatgt	7200
gaactgggtt	acggccctgt	aggcgcagca	tccctttct	acgggtacgt	cgtatgcctg	7260
cgccgccttc	cgaggcgcagg	tgtgggtgag	cgcaaaagggt	tccctgacca	tgacttttgag	7320
gtactggat	ttgaagtcag	tgtcgccca	tccggccctgc	tcccaagagca	aaaagtccgt	7380
gcgcgttttg	gaacgcggat	ttggcagggc	gaaggtgaca	tcgttgaaga	gtatcttcc	7440
cgcgcgaggc	ataaaagtgc	gtgtgatgc	gaagggtccc	ggcacctcgg	aacgggttgg	7500
aattacctgg	gccccggcgg	cgatctcg	aaagccgtt	atgttgggc	ccacaatgt	7560
aagttccaaag	aagcgccgg	tgccttgc	ggaaggcaat	ttttaagtt	cctcgttaggt	7620
gagctttca	ggggagctga	gcccgtgtc	tgaaggggcc	cagtctgca	gatgagggtt	7680
ggaagcgacg	aatagagctcc	acaggtcac	ggcattagc	atttgcagg	ggtcgcgaaa	7740
ggtcctaaac	tggcgaccta	tggccattt	ttctgggtgt	atgcagtaga	agtaagcgg	7800
gtcttgcgttcc	caggcggtccc	atccaagtt	cgccgtttag	tctcgccgg	cagtcaactag	7860
aggctcatcc	ccggccgaact	tcatgacc	catgaagggc	acgagctgt	tccaaaggc	7920
ccccatccaa	gtatagggtct	ctacatcgta	ggtgacaaag	agacgctcg	tgcgaggatg	7980
cgagccgatc	ggggaaagaact	ggatctcccg	ccaccaattt	gaggagttgc	tattgtatgt	8040
gtgaaagtag	aagttccctgc	gacggccga	acactcgtc	tggctttgt	aaaaacgtgc	8100
gcagacttgg	caggcggtca	cgggctgtac	atctgcac	aggttgac	gacgaccgc	8160
cacaaggaaag	cagagtggga	atttgagccc	ctcgccgtgg	gggtttggct	ggtggcttc	8220
tacttcggct	gcttgcctt	gaccgtctgg	ctgtcgagg	ggagttacgg	tggatcgac	8280
caccacgccc	cgcgagccca	aagtccagat	gtccgcgcgc	ggcggtcgga	gcttgcgtac	8340

aacatcgcg	agatgggagc	tgtccatgg	ctggagctcc	cgcggcgta	ggtcaggcgg	8400
gagcttgc	aggtttaccc	cgcataagac	ggtcaggcgg	cgggctagat	ccaggtata	8460
cctaattcc	aggggctgg	tggtggcg	gtcgatgg	tgcaagagc	cgcattcccg	8520
cggcgcact	acggtaccgc	ggggggggcg	gtggggcccg	gggggtgtcc	tggatgtgc	8580
atctaaaagc	ggtgacgcgg	ggagggccccc	ggaggttagg	ggggctccgg	acccgcggg	8640
agagggggca	ggggcacgtc	ggcgccgc	gccccaggg	gctgggtgt	cgcgcgtagg	8700
ttgctggcga	acgcgacgac	ggggcggtt	atctcttgc	tctggcgct	ctgcgtgaag	8760
acgacgggcc	cggtgagctt	gagcctgaaa	gagagttcg	cagaatca	ttcggtgtcg	8820
ttgacggcgg	cctggcgaa	aatctctgc	acgtctctg	agttgtctt	ataggcgatc	8880
tcggccatga	actgctcgat	ctttcttcc	tggagatctc	cgcgtccgg	tcgcgtccacg	8940
gtggggcg	ggtcggttgg	aatgcgggg	atgagctgc	agaaggcg	gaggcctccc	9000
tcgttccaga	cgcggctgt	gaccacgccc	ccttcggcat	cgcggcg	catgaccacc	9060
tcgcgcgat	tgagctccac	gtgcggcg	aagacggcg	agttcgcag	ggtcgaaag	9120
aggtatgt	gggtgggtgg	ggtgttct	gccaacaga	agtacataac	ccagcgatc	9180
aacgtggatt	cgttataatc	ccccaaaggc	tcaaggcg	ccatggctc	gtagaaagtcc	9240
acggcgaagt	taaaaactg	ggagttcg	gccccacagg	ttaactcc	ctccagaaga	9300
cgatgagct	cgcgacagt	gtcgccacc	tcgcgtca	aggctacagg	ggcctttct	9360
tcttcttca	tctctcttc	cataaggggc	tccttctt	tttcttctt	cggcggttgg	9420
ggagggggg	cacggcg	acgacggcg	acccgggg	ggtcacaaa	ggtcgatc	9480
atctcccg	ggcgacggcg	catggtctc	gtgacggcg	ggcgttctc	gcggggggcg	9540
agttggaga	cgccgcccgt	catgtcccg	ttatgggtt	gcgggggg	gccatcgcc	9600
aggatacgg	cgctaacat	gcatcttca	aattgttgc	tagtactcc	gcccggagg	9660
gacctgag	agtccgcatt	gaccggatc	gaaaacctt	cgagaaaggc	gtcttaaccag	9720
tcacagtgc	aaaggatgt	gaccatgt	ggggcg	ggggcg	gtcggttgg	9780
tttctggcg	aggtgctgt	gatgtat	ttaaagtgg	cgttcttgc	acggcgatg	9840
gtcgacagaa	gcaccatgt	tttgggtcc	gcctgtc	tgccgcgg	gtcgccatg	9900
ccccaggctt	cgttttgaca	tcggcg	tcttgtat	agtcttgc	gacccttct	9960
accggca	cttcttctt	ttccttctt	cctgcatttc	ttgcatttat	cgctcgccg	10020
gccccgg	ttggccgt	gtggcc	tttcttcc	tcgtgt	cccgaaagccc	10080
ctcatcg	gaaggcagg	tagtccg	acaacgc	cgcttaat	ggcctgtgc	10140
acctcg	gggttagact	gaagtcat	atgtcc	agcgggtt	tgccgcgt	10200
ttgatgtt	aagtgcagg	ggccat	gaccagg	cgttcttgc	acccggctgc	10260
gagagctcg	tgtacctg	acgcg	gccc	caaatacgt	gtcggtgca	10320
gtccgcacca	ggta	actgtt	ttcc	gtcggttgc	gtagaggggc	10380
cagcttaggg	ttggcgggg	tccgggg	agat	ccggcttgc	atgatatcc	10440
tagatgtacc	ttggacat	gggtat	ccgg	ttggaggc	cgaaagtgc	10500
cgacgcgg	tccagat	g	ccgg	ccatggc	gacgc	10560
ccggc	ggcgc	ggcgc	aaaa	ggcgttgc	ctgtaa	10620
ggcactt	ctgtt	gttgc	ac	aaaaggag	ccgtaa	10680
tcgagcccc	tatccgg	tccgg	ccatgc	tacggccc	gtgtcg	10740
cagg	gtcg	acgt	cgaca	gtcttctt	ccgcgg	10800
gtcg	cg	gggg	gggt	gttgc	aggcg	10860
gcgaa	aggat	taat	ggct	gggtt	gggttgc	10920
gcggg	acccc	gttgc	gggg	tat	gggttgc	10980
ccgt	catg	gggt	gggt	tttcc	gggttgc	11040
ttttcc	caga	tttgc	gggg	ccctt	gggttgc	11100
caag	gac	ggc	gggg	tcctt	gggttgc	11160
acat	ccgc	ttgc	gggg	tcacc	gggttgc	11220
ca	actt	ggg	gggg	cccg	gggttgc	11280
cggt	accc	gggt	gggg	ccgg	gggttgc	11340
ctgtt	cg	acc	gggg	ccgg	gggttgc	11400
gggcgc	gc	ggc	gggg	ccgg	gggttgc	11460
cccgac	gc	ggc	gggg	ccgg	gggttgc	11520
accgcata	cg	ggc	gggg	ccgg	gggttgc	11580
gtgcgtac	gc	ggc	gggg	ccgg	gggttgc	11640

gtaagcgcgc	tggagcaaaa	cccaaatacg	aagccgcgtca	tggcgcagct	gttccttata	11700
gtgcagcaca	gcagggacaa	cgaggcatc	agggatgcgc	tgctaaacat	agtagagccc	11760
gagggccgt	ggctgctcg	tttgataaaac	atcctgcaga	gcatagtgtt	gcaggagcgc	11820
agcttggcc	tggctgacaa	ggtggccgc	atcaactatt	ccatgcttag	cctgggcaag	11880
ttttacgccc	gcaagatata	ccataccct	tacgttccca	tagacaagga	ggtaaaagatc	11940
gaggggttct	acatgcgc	ggcgctgaag	gtgcttacct	tgagcgcacg	cctgggcgtt	12000
tatcgcaacg	agcgcatcca	caaggccgt	agcgtgagcc	ggccgcgcga	gctcagcgcac	12060
cgcgagctga	tgcacagcct	gcaaaggccc	ctggctggca	cgggcagcgg	cgatagagag	12120
gcccagtcct	actttgacgc	gggcgctgac	ctgcgctggg	ccccaaggcg	acgcgcctcg	12180
gaggcagctg	ggggccggacc	tgggctggcg	gtggcaccgg	cgcgcgtgg	caacgtcgcc	12240
ggcgtggagg	aatatgacga	ggacgatgag	tacgagccag	aggacggcga	gtactaagcg	12300
gtgatgttct	tgatcagatg	atgcaagacg	caacggacc	ggcgggtgcgg	gcccgcgtgc	12360
agagccagcc	gtccggcctt	aactccacgg	acgactggcg	ccaggtcatg	gaccgcata	12420
tgtcgctgac	tgcgcgaat	cctgacgcgt	tccggcagca	gccgcagcc	aaccggctct	12480
ccgcatttct	ggaagegggt	gtccggcgc	gcgcääacc	cacgcacag	aagggtgcgtt	12540
cgatcgtaaa	cgcgctggcc	gaaaacaggg	ccatccggcc	cgacgaggcc	ggcctggctt	12600
acgacgcgt	gcttcagcgc	gtggctcggt	acaacagccg	caacgtcgag	accaacctgg	12660
accggctgg	ggggatgtg	cgcgaggccg	tggcgagcg	tgagcgcgcg	cagcagcagg	12720
gcaacctggg	ctccatggtt	gcactaaacg	ccttcctgag	tacacagcc	gccaacgtgc	12780
cgccgggaca	ggaggactac	accaacttt	tgagcgcact	gcccctaattg	gtgactgaga	12840
caccgcaaaag	tgaggtgtac	cagtctggc	cagactattt	tttccagacc	agtagacaag	12900
gcctgcagac	cgtaaacctg	agccaggctt	tcaaaaactt	gcagggctg	tgggggtgc	12960
gggctccac	aggcgaccgc	gcgaccgtgt	ctagcttgc	gacgcccac	tgcgcctgt	13020
tgctgctgt	aatacgcccc	ttcacggaca	gtggcagcgt	gtccgggac	acatacctag	13080
gtcacttgc	gacactgtac	cgcgaggcca	tagtcaggc	gcatgtggac	gagcatactt	13140
tccaggagat	tacaagtgtc	agccgcgcgc	tggggcagga	ggacacggc	agctggagg	13200
caaccctaaa	ctacctgtg	accaaccggc	ggcagaagat	cccctcggt	cacagttaa	13260
acagcgagga	ggagcgcatt	ttgcgctacg	tgcagcagag	cgtgagcctt	aacctgtatgc	13320
gcgacggggt	aaaccccagc	gtggcgctgg	acatgaccgc	gchgacacatg	gaaaccggca	13380
tgtatgcctc	aaaccggccg	tttatcaacc	gcctaatagg	ctacttgcat	cgcgcggccg	13440
ccgtgaaccc	cgagtattt	accaatgcca	tcttgaaccc	gcactggcta	ccgcggccctg	13500
gtttctacac	cgggggattc	gaggtgccc	aggtaacga	tggattcctc	tggacgaca	13560
tagacgacag	cgtgttttcc	ccgcaaccgc	agaccctgt	agagttgcaa	cagcgcgagc	13620
aggcagaggg	ggcgcgtgcg	aaggaaagct	tccgcaggcc	aagcagctt	tccgatctag	13680
gcgctgcggc	cccgccgtca	gatgcttagt	gcccatttcc	aagcttgata	gggtcttta	13740
ccagcaetcg	caccacccgc	ccgcgcctgc	tgggcagga	ggagttacta	aacaactcg	13800
tgctgcagcc	gcagcgcgaa	aaaaacctgc	tcccgccatt	tcccaacac	gggatagaga	13860
gccttagtgg	caagatgagt	agatggaaa	cgtacgcgc	ggagcacagg	gacgtgcgg	13920
gccccggccc	gcccacccgt	cgtcaaaggc	acgaccgtca	gcgggggtctg	gtgtggggagg	13980
acgatgactc	ggcagacgac	agcagcgtcc	tggattttgg	aggaggtggc	aaccctttt	14040
cgcacccctcg	ccccaggctg	ggggagaatgt	ttttaaaaaaa	aaaaagcatg	atgaaaata	14100
aaaaactcac	caaggccatg	gcaccggagc	ttgtttttt	tgtattcccc	ttagtatgcg	14160
gcgcgccccg	atgtatgagg	aaggtcctcc	tccctctac	gagagtgtgg	tgagcgcggc	14220
gccagttggcg	gcccggctgg	gttctccctt	cgatgctccc	ctggaccgc	cggttgcgtt	14280
tccgcggta	ctgcggccta	ccggggggag	aaacagcata	cgttactctg	agttggcacc	14340
cctattcgac	accacccgt	tgtacctgtt	ggacaacaag	tcaacggatg	tggcatcccc	14400
gaactaccag	aacgaccaca	gcaactttt	gaccacggc	attcaaaaaca	atgactacag	14460
cccgggggag	gcaagcacac	agaccatcaa	tcttgacgc	cggtcgca	ggggcggcga	14520
cctgaaaacc	atcctgcata	ccaacatgcc	aaatgtgaac	gagttcatgt	ttaccaataa	14580
gttttaaggcg	cgggtgtatgg	tgtcgctt	gcctactaa	gacaatcagg	tggagctgaa	14640
atacgagtgg	gtggagttca	cgctgccc	gggcaactac	tccgagacca	tgaccataga	14700
ccttatgaaac	aacgcgatcg	tggagacta	cttggaaatgt	ggcagacaga	acgggggttct	14760
ggaaagcgac	atcggggtaa	agtttgcac	ccgcaacttc	agactgggt	ttgacccctg	14820
caactggctt	gtcatgcctg	gggtatatac	aaacgaagcc	ttccatccag	acatcatttt	14880
gctgccagga	tgcggggtgg	acttcaccca	cagccgcctg	agcaacttgt	tggcatccg	14940

caagcggcaa	cccttcagg	agggcttag	gatcacctac	gatgatctgg	aggggtggtaa	15000
cattcccga	ctgttggatg	tggacgccta	ccagggcagc	ttgaaagatg	acacccaa	15060
ggcggggg	ggcgaggcg	gcagcaacag	cagtggcagc	ggcgccgaa	agaactcaa	15120
cgcggcagcc	gcccgaatgc	agccgttga	ggacatgaac	gatcatgcca	ttcgccgg	15180
cacccggcc	acacgggctg	aggagaagcg	cgctgaggcc	gaagcagccg	ccgaagctgc	15240
cgcccccgct	gcccgaaccccg	aggtcgagaa	gcctcagaag	aaaccggta	tcaaaacccct	15300
gacagaggac	agcaagaaac	gcagttacaa	cctaataa	aatgacagca	ccttcaccca	15360
gtaccgcagc	tgttaccc	catacaacta	cgccgaccct	cagaccggaa	tccgctcatg	15420
gaccctgctt	tgcaactc	acgtaacctg	cggtcggag	caggtctact	ggtcgttgc	15480
agacatgtat	caagaccccg	tgaccc	ctccacgcgc	cagatcagca	actttcccg	15540
ggtgggcgc	gagctgttgc	ccgtgcact	caagagctt	tacaacgacc	aggccgtcta	15600
ctcccaactc	atccggcc	ttaccc	gaccacgt	ttcaatcgt	ttcccggagaa	15660
ccagatttt	gcgcggcc	cagccccac	catacacc	gtcagtgaaa	acgttccctgc	15720
tctcacagat	cacgggacgc	taccgc	caacagcatc	ggaggat	acgcgagtac	15780
cattactgac	gcccacgc	gcaccc	ctacgttac	aaggccctgg	gcatagtctc	15840
gcccgcgtc	ctatcgac	gcactt	agaagacatg	tccatccta	tatcgcccag	15900
caataacaca	ggctggggcc	tgcg	tttggcgggg	ccaagaagcg	15960	
ctccgaccaa	caccc	acgtgc	gcgtgcgg	gcccctgg	gcccgcacaa	16020
acgcggccgc	actgggc	ccaccgtc	tgacgc	gacgcgg	tgaggagg	16080
gcccacactac	acgcccacgc	cgccacc	gtccac	gacgcgg	ttcagacc	16140
ggtgcggg	gcccggcg	atgt	aaat	cgaggc	tagcacgt	16200
ccaccgcgc	cgacccggc	ctgccc	acgcgcgg	gcccctgc	ttaacccgc	16260
acgtgc	ggccgacgg	cggccat	ggccgc	aggctgg	cggttattgt	16320
cactgtgc	cccagg	ttca	ggcgc	ccat	gtc	16380
tatgactc	ggtgc	gcaac	gtgt	gactcg	ggccgc	16440
cgtcccgt	cgcac	cccc	cgca	ctagatt	agaaaaaa	16500
gtactgtt	atgtatcc	cg	gcgc	gtatgt	acttagact	16560
caaagaagag	atgtcc	tc	ggac	ggatctat	ggcccc	16620
gcaggattac	aaggcc	agct	aaagcg	gtcaaaa	agaaggaa	16680
tgaacttgc	gacg	agggt	actgc	cgctacc	atgtatgt	16740
gaaagg	tcgtaaa	at	ccgg	accgt	gggtac	16800
tgagc	cc	cc	gcac	ct	ttac	16860
gttt	acc	aca	agcg	gtat	gtc	16920
gttt	gca	ac	ggac	gtt	ggat	16980
gttt	cc	cc	gtt	ccaa	acta	17040
gcagg	cc	cc	gtt	cc	cc	17100
tgact	cc	cc	gtt	cc	cc	17160
ggaaaaat	acc	tt	ggc	cc	cc	17220
gg	gg	gg	gg	cc	cc	17280
gg	gg	gg	gg	cc	cc	17340
gg	gg	gg	gg	cc	cc	17400
gg	gg	gg	gg	cc	cc	17460
gg	gg	gg	gg	cc	cc	17520
gg	gg	gg	gg	cc	cc	17580
gg	gg	gg	gg	cc	cc	17640
gg	gg	gg	gg	cc	cc	17700
gg	gg	gg	gg	cc	cc	17760
gg	gg	gg	gg	cc	cc	17820
gg	gg	gg	gg	cc	cc	17880
gg	gg	gg	gg	cc	cc	17940
gg	gg	gg	gg	cc	cc	18000
gg	gg	gg	gg	cc	cc	18060
gg	gg	gg	gg	cc	cc	18120
gg	gg	gg	gg	cc	cc	18180
gg	gg	gg	gg	cc	cc	18240

tggtagatgg	cctggcctct	ggcattagcg	gggtgggtgga	cctggccaac	caggcagtgc	18300
aaaataagat	taacagtaag	cttgatcccc	gcccctccgt	agaggagct	ccaccggccg	18360
tggagacagt	gtctccagag	gggcgtggcg	aaaacgtcc	gcccggccac	agggaaagaaa	18420
ctctggtgac	gcaaatagac	gagcctccct	cgtacgagga	ggcactaaag	caaggcctgc	18480
ccaccacccg	tccatcgcg	cccatggcta	ccggaggatgt	gggcccagac	acacccgtaa	18540
cgctggacct	gcctcccccc	gcccacaccc	agcagaaaacc	tgtgctgcca	ggcccgaccg	18600
ccgttggtgt	aaccgtctt	agccgcgcgt	ccctgcgcgg	cgccggccagc	ggtccgcgtat	18660
cgttgcggcc	cgtagccagt	ggcaactggc	aaagcacact	gaacagcatc	gtgggtctgg	18720
gggtgcaatc	cctgaagcgc	cgacgatgt	tctgaatagc	taacgtgtcg	tatgtgtgtc	18780
atgtatgcgt	ccatgtcgcc	gccagaggag	ctgctgagcc	gcccgcgcgc	cgctttccaa	18840
gatggctacc	ccttcgtatga	tgccgcagtg	gtcttacatg	cacatctcg	gccaggacgc	18900
ctcggagttac	ctgagccccg	ggctggtgca	gtttgcgcgc	gcccggaga	cgtacttcag	18960
cctgaataac	aagggttagaa	acccacccgt	ggcgcctact	cacgacgtg	ccacagacccg	19020
gtccccagct	ttgacgctgc	ggttcatccc	tgtggaccgt	gaggatactg	cgtactcgta	19080
caaggcgcgg	ttaacccttag	ctgtgggtga	taaccgtgtg	ctggacatgg	cttccacgt	19140
ctttgacatc	cgccgcgtgc	tggacagggg	ccctactttt	aaggccctact	ctggcactgc	19200
ctacaacgccc	ctggctccca	agggtgcggc	aaatcctgc	aatgggatg	aagctgtac	19260
tgctcttgaa	ataaaacctag	aagaagagga	cgatgacaac	gaagacgaag	tagacgagca	19320
agctgagcag	caaaaaactc	acgtattttg	gcaggcgcct	tattctgtta	taaatattac	19380
aaaggagggt	attcaaataag	gtgtcgaagg	tcaaacacct	aaatatgccg	ataaaaacatt	19440
tcaacctgaa	cctcaaataag	gagaatctca	gtggtacgaa	actgaaatata	atcatgcgc	19500
tgggagagtc	ctaaaaaaga	ctaccccaat	gaaaccatgt	tacggttcat	atgaaaaacc	19560
cacaatgaa	aatggagggc	aaggcattct	tgtaaagcaa	caaaatgaa	agctagaaaag	19620
tcaagtggaa	atgcaatttt	tctcaactac	tgaggcgacc	gcaggcaatg	gtgataactt	19680
gactcctaaa	gtggttatgt	acagtgaaga	tgtagatata	gaaaccccg	acactcatat	19740
ttcttacatg	ccactattata	aggaaggtaa	ctcagcagaa	ctaattggcc	aaacatcttat	19800
gccaacagg	cctaattaca	ttgcttttag	ggacaatttt	attggctaa	tgtattacaa	19860
cagcacgggt	aatatgggt	ttctgggggg	ccaagcatcg	cagttgaatg	ctgtttaga	19920
tttgcagac	agaaacacag	agctttcata	ccagctttt	cttgatttca	ttgggtatag	19980
aaccaggta	ttttctatgt	ggaatcaggc	tgttgacagc	tatgatccag	atgttagaat	20040
tattgaaaat	catggaactg	aagatgaact	tccaaattac	tgctttccac	tgggaggtgt	20100
gattaataca	gagactctt	ccaaggtaaa	acctaaaaca	gttcaggaaa	atggatggga	20160
aaaagatgt	acagaatttt	cagataaaaa	tgaataaaga	gttggaaata	attttgcct	20220
ggaaatcaat	ctaaatgcct	acctgtggag	aaatttctgt	tactccaaca	tagcgtgt	20280
tttgcggc	aagctaaagt	acagtccttc	caacgtaaaa	atttctgata	acccaaacac	20340
ctacgactac	atgaaacaatgc	gagtgggtc	tcccgggtt	gtggactgt	acattaaacct	20400
tggagcacgc	tggccctt	actatatgg	caacgtcaac	ccatttaacc	accacccgca	20460
tgtggccgt	cgctaccgc	caatgttgc	ggcaatgtt	cgctatgtgc	ccttccacat	20520
ccaggtgcct	cagaaggttc	ttggccat	aaacccctt	ctccgcgg	gctatacac	20580
ctacgagtgg	aacttcagga	aggatgttaa	catggttctg	cagagctccc	tagaaatga	20640
cctaagggtt	gacggagcca	gcattaagtt	tgtatgcatt	tgcccttacg	ccacccctt	20700
ccccatggcc	cacaacaccc	cctccacgc	tgaggccatg	cttagaaacg	acaccaacga	20760
ccagtcctt	aacgactatac	tctccgcgc	caacatgctc	taccctatac	ccgccaacgc	20820
taccaacgtg	cccatatcca	tcccccctcc	caactggcg	gctttccgcg	gctgggcctt	20880
cacgcgcctt	aagactaagg	aaaccccatc	actgggctcg	ggctacgacc	cttattacac	20940
ctactctggc	tctataccct	acctagatgg	aaccttttac	ctcaaccaca	cctttaagaa	21000
ggtggccatt	acccttgact	tttctgtcag	ctggcctggc	aatgaccgc	tgcttacccc	21060
caacgagtt	gaaattaagc	gctcagttga	cggggagggt	tacaacgtt	cccagtgtaa	21120
catgacccaaa	gactggttcc	ttgtacaaat	gctagctaac	tacaacattg	gttaccagg	21180
cttctatatac	ccagagagct	acaaggaccc	catgtactcc	ttttagtataa	actccagcc	21240
catgagccgt	caggtgggtgg	atgataactaa	atacaaggac	taccaacagg	tggcatactt	21300
acaccaacac	aacaactctg	gattttgtgg	ctacccgtcc	cccaccatgc	gcaaggaca	21360
ggcctaccct	gctaacttcc	cctatccgt	tataggcaag	accgcagg	acagcattac	21420
ccagaaaaaa	tttctttgcg	atgcaccctt	ttggcgcatac	ccattcttca	gtactttat	21480
gtccatgggc	gcactcacag	acctgggcca	aaacccctc	tacgccaact	ccgcccacgc	21540

gctagacatg acttttgagg tggatccat ggacgagccc acccttctt atgtttgtt	21600
tgaagtctt gacgtggtcc gtgtgcaccc gcccacccgc ggcgtcatcg aaaccgtgta	21660
cctgcgcacg cccttctcgg cccgcaacgc cacaacataa aagaagcaagc aacatcaaca	21720
acagctgccc ccatgggctc cagtgagcag gaactgaaag ccattgtca agatcttgg	21780
tgtgggcccattttttggg cacctatgac aagcgtttc caggcttgg ttctccacac	21840
aagctctgcct ggcctatgt caatacggcc ggtcgcgaga ctgggggcgt acactggatg	21900
gccttgcct ggaacccgc a ctc aaaaaca tgta cctt ct tgagccct tggctttct	21960
gaccagcgcac tcaagcagg tta cca ggtt gat gat cacttgcg cccgtacg	22020
attgcttctt ccccccaccc ctgtataacg ctggaaaagt ccacccaaag cgtacagg	22080
cccaactcgg cccctgtgg actatttcgc tgcatgttcc tccacgcctt tgccaactgg	22140
ccccaaactc ccatggatca caacccaccatgaaacccat tta cccgggtt acccaactcc	22200
atgctcaaca gtc cccagg tca gcccacc ctgcgtcgca accaggaaca gctctacgc	22260
ttcctggagc gccactcgcc ctacttccgc a gcccacgtg cccagattt gaggccact	22320
tcttttgc acttggaaaaa catgtaaaaa taatgtacta gagaacactt caataaaggc	22380
aaatgtttt atttgtacat tctgggtga ttat tacc ccaccccttgc cgtctgcgc	22440
gtttaaaaat caaagggtt ctgcgcgc a tcgctatgcg ccactggcag ggacacgttgc	22500
cgatactgg ttttagtgc ctactttaac tca gggcaca ccatccgc cagctcggt	22560
aagtttcaac tccacaggct ggcacccatc acca acgcgt ttagcagg tggccgcgat	22620
atcttgaagt cgcgttggg gctccgccc tgcgcgcg agttgcata cacagggttgc	22680
cagcactgga acactatcag cccgggttgc tgcacgcgtt ccagcacgtt ctgtcggag	22740
atcagatccg cgtccagg tccgcgttgc ctca gggcga acggagtcaaa ctttggtagc	22800
tgcctccca aaaaggcgc gtc cccaggc tttagtgc actgcaccc tagtgcac	22860
aaaagggtgac cgtgcgcgtt ctgggcgtt gat gacccgc cctgcataaa agcctgtac	22920
tgctttaaaag ccacccgtac tttgcgcctt ctagagaaca acatccgc a gacttgcgc	22980
aaaaactgtat tggccgcata ggcgcgtcg tgcacgcgc accttgcgtt ggttggag	23040
atctgcacca catttgcgc ccacccgttca tca cgcgttctt gggcttgc agactgctcc	23100
ttcagcgcgc gtc cccgtt ttcgcgtc acatccattt caatcacgtg ctccttattt	23160
atcataatgc ttccgttagt acacttaagc tgcgccttca ttcgcgcgca gcccgtcagc	23220
cacaacgcgc agcccggttgc ttcgtatgc tttaggtca cctctgc aaaa c gactgcagg	23280
tacgcctgca ggaatgc cccatcatgc acaaagg tctt ggttgcgtt gaagg tgcagc	23340
tgcaacccgc ggtgcctc gttcaggcag gtcttgcata cggccgcag agcttccact	23400
ttggcaggca gtagttgaa gttcgcctt agatcgat tccacgttgc cttgttccatc	23460
agcgcgcgc cagccctccat gccccttcc cccgcgcac cgcgcgcac actcagcgg	23520
ttcatac cccat tttctact ttcgcgttgc tttaggtca cctctgc aaaa c gactgcagg	23580
ataccacgcgc ccactgggtc gtcttcatcc acgcgcgc acctcttttgcgtt acctcttgc	23640
ccatgcgttgc tttagcaccgg tgggttgc tttgtgc cccat ttttgc cccat ttttgc	23700
ctttcttcc cgcgttccac gattacccctt ggtgttgc ggcgcgcgg cttggagaa	23760
gggcgcgttct tttcttccat gggcgcata gccaaatccg cccgcgcagg cgcgcgc	23820
gggcgttgc tgcgcgcac cagcgcgttgc tttgtatgcgtt cttcgcgtc ctcgcactcg	23880
atacgcgc tcatccgc ttttggggc gcccggggag gcccggcga cggggacgg	23940
gacgacacgt ctcctcatgtt tggggacgt cgcgcgcac cgcgcgcgc cttgggggtg	24000
gtttcgctc gtc cccacttcc cccactggcc atttcttctt cctataggca gaaaaagatc	24060
atggagttagtgc tggagaagaa ggacaggcata accggccccc ctgatgc caccacccgc	24120
tccaccatgc ccccaacgc gcctaccacc ttccgcgtc aggcacccccc gtttgcagg	24180
gaggaagtga ttatcgagca ggacccagg tttgtaa gca gacgcacca gggccgtca	24240
gtaccaacag aggataaaaaa gcaagacccag gacaacgcac aggcacccacca ggaacaagtc	24300
gggggggggg acgaaaggca tggcactac ctagatgtgg gagacgcacgt gctgttgaag	24360
catctgcacgc gccagg tgcgc cattatctgc gacgcgttgc aagagcgcac cgcgttgc	24420
ctcgccatag cggatgtca cttgccttac gacgcgcacc tattctcacc ggcgcgttcc	24480
cccaacgcgc aaaaaacggc cccatgcgc gcccacccgc gcctcaaccc tca cccgcgt	24540
tttgcgtgc cagagg tgc tccacccat cccat ttttccat cccat ttttccat cccat	24600
ctatctgcgc tggccaaacgc cccgcgcgc gacaacgcac tggcccttgcgc gcaggccgc	24660
gtcatactgc ttttgcgttcc gtc aacggaa gtc cccaaaatccat ttttgcagg ttttgcagg	24720
gacgagaagc ggcggc aaaa c cgcgttgc a cccat ttttgcagg ttttgcagg	24780
ggagtgttgg tggaaactcga gggtgc aac ggcgcctt ggcgtactaaa acgcacgc	24840

gaggtcaccc	actttgccta	cccggcac	tttacccatccc	ccaagg	tcat	gagcac	agtc	24900
atgagtgagc	tgatcg	tcg	ccgtgcgc	ccgc	gggat	gc	tttgc	24960
caaacagagg	agggc	tacc	cg	cc	tttgc	aaa	aa	25020
cgcgagc	c	ctg	g	cc	tttgc	tttgc	aa	25080
gtggagc	tt	tg	g	cc	tttgc	tttgc	aa	25140
aaaacat	tc	ac	ac	cc	tttgc	tttgc	aa	25200
gtggagc	tc	ac	ac	cc	tttgc	tttgc	aa	25260
aaacgt	tc	tc	tc	cc	tttgc	tttgc	aa	25320
tacttat	tt	ta	tg	cc	tttgc	tttgc	aa	25380
gagtgc	aa	cc	cc	cc	tttgc	tttgc	aa	25440
gccttca	ac	cg	cc	cc	tttgc	tttgc	aa	25500
cttaaa	ac	cc	cc	cc	tttgc	tttgc	aa	25560
aggaactt	ta	cc	cc	cc	tttgc	tttgc	aa	25620
gacttt	tg	cc	cc	cc	tttgc	tttgc	aa	25680
tgc	ca	cc	cc	cc	tttgc	tttgc	aa	25740
gttctact	gg	tt	tt	tt	tttgc	tttgc	aa	25800
aattc	cg	cc	cc	cc	tttgc	tttgc	aa	25860
cctg	ac	gg	gg	gg	tttgc	tttgc	aa	25920
tac	tt	gg	gg	gg	tttgc	tttgc	aa	25980
caatccc	cc	cc	cc	cc	tttgc	tttgc	aa	26040
ggccaa	tt	cc	cc	cc	tttgc	tttgc	aa	26100
gttta	cc	cc	cc	cc	tttgc	tttgc	aa	26160
tatc	ag	cc	cc	cc	tttgc	tttgc	aa	26220
gccc	cc	cc	cc	cc	tttgc	tttgc	aa	26280
cgagg	gg	gg	gg	gg	tttgc	tttgc	aa	26340
cgaag	gg	gg	gg	gg	tttgc	tttgc	aa	26400
gtt	tt	cc	cc	cc	tttgc	tttgc	aa	26460
tc	at	cc	cc	cc	tttgc	tttgc	aa	26520
tc	at	cc	cc	cc	tttgc	tttgc	aa	26580
cc	gt	cc	cc	cc	tttgc	tttgc	aa	26640
ct	cc	cc	cc	cc	tttgc	tttgc	aa	26700
tc	at	cc	cc	cc	tttgc	tttgc	aa	26760
tc	at	cc	cc	cc	tttgc	tttgc	aa	26820
cc	gg	gg	gg	gg	tttgc	tttgc	aa	26880
ag	tt	tt	tt	tt	tttgc	tttgc	aa	26940
ctt	at	cc	cc	cc	tttgc	tttgc	aa	27000
at	gg	gg	gg	gg	tttgc	tttgc	aa	27060
gg	cc	cc	cc	cc	tttgc	tttgc	aa	27120
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27180
at	gg	gg	gg	gg	tttgc	tttgc	aa	27240
gg	cc	cc	cc	cc	tttgc	tttgc	aa	27300
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27360
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27420
gg	cc	cc	cc	cc	tttgc	tttgc	aa	27480
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27540
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27600
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27660
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27720
tt	gg	gg	gg	gg	tttgc	tttgc	aa	27780
gg	cc	cc	cc	cc	tttgc	tttgc	aa	27840
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27900
cc	cc	cc	cc	cc	tttgc	tttgc	aa	27960
tt	gg	gg	gg	gg	tttgc	tttgc	aa	28020
tt	gg	gg	gg	gg	tttgc	tttgc	aa	28080
gacc	cc	cc	cc	cc	tttgc	tttgc	aa	28140

gttgcacatct ctgtgctgag tataataaaat acagaaaatta aataatactg gggctccat	28200
cgccatctg taaaacgcac cgtcttccacc cgcggcaagca aaccaaggcg aaccccttacct	28260
ggtacttttta acatctctcc ctctgtgatt tacaacagt tcaacccaga cggagtgagt	28320
ctacgagaga acctctccga gtcagctac tccatcagaa aaaacaccac cctccttacc	28380
tgccgggaac gtacgagtgc gtcacccggcc gctgcaccac acctaccgc tgaccgtaaa	28440
ccagactttt tccggacaga cctcaataac tctgtttacc agaacaggag gtgagcttag	28500
aaaaccctta gggattttagg ccaaaggcgc agctactgtg gggtttatga acaattcaag	28560
caactctacg ggctattctta attcaggttt ctctagaatc ggggttgggg ttattctctg	28620
tcttgcatttcttactt ttataactaac gcttctctgc ctaaggctcg cgcctctgt	28680
tgtgcacatt tgcatttttattt gtcagctttt taaacgcgtt ggtgcggccacc caagatgatt	28740
aggtacataa tccctaggtt actcaccctt gctcagcccc acggattacc caaaagggtg	28800
gattttaagg agccggctt taatgttaca ttgcagctg aagctaatga gtgcaccact	28860
cttataaaat gcaccacaga acatggaaatg ctgcttattt gcccacaaaaaaa caaaatttggc	28920
aagtatgtg tttatgttat ttggcagcca ggtgacacta cagagtataa tggttacagt	28980
ttccagggtt aaagtctataa aactttatg tatacttttccat ttttatga aatgtgcgac	29040
attaccatgt acatgagcaa acagtataag ttgtggcccc cacaatttgg tttggaaaac	29100
actggcactt tctgctgcac tgctatgcta attacagtgc tcgcttggg ctgtaccctta	29160
ctctatatta aataaaaaat gagacgcgc tttatttggg aaaaagaaaaat gccttaattt	29220
actaagttac aaagcttaatg tcaccactaa ctgctttact cgctgttgc aaaaacaaattt	29280
aaaaaaatgttta gcattataat tagaaatggg tttaaaccccccc cgggttcatttt cctgctcaat	29340
accattcccc tgaacaaatgg actctatgtt ggtatgttc cagcgctaca accttgaagt	29400
cagggttctt ggtatgtcgc atctgactttt gggcagcacc tgccggccg atttggccca	29460
gttccaaatc acgcggccac cctaaacagag atgaccaaca caaccaacgc gcccggcgt	29520
accggactta catctaccac aaatacacc ccaagtttctg cttttgtcaa taactggat	29580
aacttggca tttttttttt ctccatagcg cttatgtttt tatgtccttat tattatgtgg	29640
ctcatctgttgc gccttaaagcg caaacgcgc cggccacccca tctatagtcc catattgtg	29700
ctacacccaa acaatgttgg aatccataga ttggacggac tggaaacacat gttttttct	29760
cttacagtat gattaaatgtt gacatgattt ctcgagttt tattttactg accttgggt	29820
cgcttttttgc tgcgtgtcc acattttgtt cgggttctca catcgaaatg gactgttcc	29880
cagcccttac agtctattt ctttacggat ttgtccacccat cagcgttccatc tgccgttca	29940
tcactgtgtt catcgctt atccatgttca ttgactgggtt ctgtgtgcg tttgcataatc	30000
tcagacacca tccccatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc	30060
tatgaaatttt actgttactt ttctgttgc ttttttttttgc ttttttttttgc ttttttttttgc	30120
gaccttcaag ccttcaaaatgg atatatcatg cagatttactt ctttttttttgc ttttttttttgc	30180
tttgcataat gaaaaaaatgg atctttccatggggccat ttttttttttgc ttttttttttgc	30240
ggtgttctgc agtaccatct tagccatgttccatggggccat ttttttttttgc ttttttttttgc	30300
acgaaatagat gccatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc	30360
agtttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30420
tgttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30480
acggaaatcatg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30540
gcatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30600
gtcttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30660
acaaatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30720
taacttcatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30780
actatcttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30840
aaaaaaaat aataaaatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc	30900
tttgcacatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	30960
aaacttctccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	31020
actatcttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	31080
gtgtatccatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	31140
gtatccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	31200
cctcttagttcaatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc	31260
gaggccggca accttacccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc	31320
aagtcaaaatgttccatggggccat ttttttttttgc ttttttttttgc ttttttttttgc	31380
gtggctgcgc cccgcacccatgttccatggggccat ttttttttttgc ttttttttttgc	31440

ccgctaaacc	tgcacgactc	caaacttagc	attgccaccc	aaggaccctt	cacagtgtca	31500
gaaggaaagc	tagccctgca	aacatcaggc	cccccacca	ccaccgatag	cagtaccctt	31560
actatcactg	cctcaccctt	tcttaactact	gccactggta	gcttggcat	tgacttggaa	31620
gagcccttt	atacacaaaa	tggaaaacta	ggactaaagt	acggggctcc	tttgcattgt	31680
acagacgacc	taaacacttt	gaccgttagca	actggtccag	gtgtactat	taataataact	31740
tccttgcata	ctaaagttac	tggagccctt	gggtttgatt	cacaaggcaa	tatgcactt	31800
aatgttagcag	gaggactaag	gattgattct	caaaacagac	gccttataact	tgatgttagt	31860
tatccgttt	atgctaaaa	ccaactaaat	ctaagactag	gacagggccc	tcttttata	31920
aactcagccc	acaacttgg	tattaactac	aacaaaggcc	tttacttgtt	tacagcttca	31980
aacaattcca	aaaagcttga	ggtaaccta	agcactgcca	aggggttggat	gtttgacgct	32040
acagccatag	ccattaatgc	aggagatggg	cttgaattt	gttcacctaa	tgcacccaaac	32100
acaaatcccc	tcaaaaacaaa	aattggccat	ggcctagaat	ttgattcaaa	caaggtatg	32160
gttccttaaac	taggaactgg	ccttagttt	gacagcacag	gtgcattac	agtaggaaac	32220
aaaaataatg	ataagctaac	tttgggacc	acaccagtc	catctcctaa	ctgttagacta	32280
aatgcagaga	aagatgtaa	actcaactt	gtcttaacaa	aatgtggcag	tcaaatactt	32340
gctacagtt	cagtttggc	tgtttaaggc	agtttggctc	caatatctgg	aacagttcaa	32400
agtgcctatc	tttataaag	atttgacaa	attggagtgc	tactaaacaa	ttcccttcctg	32460
gaccaggaaat	attggaactt	tagaaatgg	gatctactg	aaggcacagc	ctatacaaaac	32520
gctgttgat	ttatgcctaa	cctatcaact	tatccaaat	ctcacggtaa	aactgccaaa	32580
agtaacattt	tcaagtcaat	ttacttaac	ggagacaaaa	ctaaacctgt	aacactaacc	32640
attacactaa	acggtacaca	ggaaacagga	gacacaactc	caagtgcata	ctctatgtca	32700
ttttcatggg	actggctgg	ccacaactac	attaatgaaa	tatttgcac	atccctttac	32760
actttttcat	acattggcc	agaataaaga	atcgtttgc	ttatgttca	acgtgtttat	32820
ttttcaattt	cagaaaattt	caagtcattt	ttcattcagt	agtatagccc	caccaccaca	32880
tagttatac	agatcaccgt	accttaatca	aactcacaga	acccttagtat	tcaacactgccc	32940
acctccctcc	caacacacag	agtacacagt	ccttctccc	cggctggcct	taaaaagcat	33000
catacatgg	gtacacagaca	tatttcttag	tgttatattc	cacacgggtt	cctgtcgac	33060
caaacgtca	tcaagtat	taataaaactc	cccccgcagc	tcaatctaagt	tcatgtcgct	33120
gtccagctgc	tgagccacag	gctgctgtcc	aacttgcgg	tgcttaacgg	gccccgaaagg	33180
agaagtccac	gcctacatgg	gggttagagtc	ataatcgtgc	atcaggatag	ggccgtgggt	33240
ctgcagcagc	gcccgaataaa	actgctggcg	ccggcgtcc	gtcctgcagg	aatacaacat	33300
ggcagtggc	tcctcagcga	tgattcgcac	ccccgcagc	ataaggcgcc	ttgtccctcg	33360
ggcacagcag	cgcaccctga	tctcaactaa	atcagcacag	taactgcagc	acagcaccac	33420
aatatttttc	aaaatccac	agtcaaggc	gctgtatcca	aagctcatgg	cggggaccac	33480
agaacccacg	tggccatcat	accacaagcg	caggtagatt	aagtggcagc	ccctcataaa	33540
cacgctggac	ataaacattt	ccttttttgg	catgttgc	ttcaccacct	ccgggtacca	33600
tataaacctc	tgattaaaca	tggcgccatc	caccaccatc	ctaaaccagc	tggccaaaac	33660
ctgcccggcc	gttataact	gcagggaaacc	ggactgtggaa	caatgcacgt	ggagagccca	33720
ggactcgtaa	ccatggatca	tcatgctct	catgatata	atgttggc	aacacaggca	33780
cacgtgcata	cacttccctca	ggattacaag	ctcccccgc	tttagaaacca	tatcccagg	33840
aacaacccat	tcctgaatca	gcgttaatcc	caactgcag	ggaagacactc	gcacgtact	33900
cacgttgc	attgtcaaaag	tgttacattt	gggcagcagc	ggatgtatct	ccagttatgtt	33960
agcgcgggtt	tctgtctcaa	aaggaggtag	acgatccctt	ctgtacggag	tgcgcccaga	34020
caacccgagat	cgtgttggc	gtagtgtcat	gccaaatgg	acgcccggacg	tagtcatatt	34080
tcctgaagca	aaaccagggt	cgggcgtac	aaacagatct	gcgtctccgg	tctccggct	34140
tagatcgctc	tgtttagtag	ttttagtata	tccactctct	caaagcatcc	aggccccccc	34200
tggcttcggg	ttctatgtaa	actccttcat	gcggcgtcc	cctgataaaca	tccaccaccg	34260
cagaataagg	cacacccagc	caacctacac	attcggtctg	cgagtccacac	acgggaggag	34320
cgggaaagagc	tggaaagaacc	atgtttttt	tttatttca	aaagattatc	aaaaacactca	34380
aatgaagat	ctattaatgt	aacgcgtcc	cctccgggt	cgtggtaaaa	ctctacagcc	34440
aaagaacaga	taatggcatt	tgtaagatgt	tgacacatgg	cttccaaaag	gcaaaacggcc	34500
ctcacgtcca	agttggacgt	aaggcttaac	cctcagggt	gaatctcttc	tataaaccatt	34560
ccagcacctt	caaccatgcc	caaataattc	tcatctcgcc	accttctcaa	tatatctcta	34620
agcaaaatccc	gaatattaaag	tccggccatt	gtaaaaatct	gctccagagc	gcccctccacc	34680
ttcagcttca	agcagcgaat	catgatttgc	aaaattcagg	ttcctcacag	acctgtataa	34740

gattcaaaaag	cggaacatta	acaaaaatac	cgcgatcccg	taggtccctt	cgcagggcca	34800
gctgaacata	atcgtgcagg	tctgcacgga	ccagcgcggc	cactccccc	ccaggaacct	34860
tgacaaaaga	acccacactg	attatgacac	gcatactcg	agctatgtca	accagcgtag	34920
ccccgatgt	agctttgtt	catggccggc	gatataaaaat	gcaagggtct	gctcaaaaaaa	34980
tcagggaaaag	cctcgcgc	aaaagaaaagc	acatcgtagt	catgctcatg	cagataaagg	35040
caggtaaagct	ccggaccac	cacagaaaaa	gacaccattt	ttctctcaaa	catgtctgcg	35100
ggtttctgca	taaacacaaa	ataaaataac	aaaaaaaacat	ttaaacatta	gaaggctgtc	35160
ttacaacagg	aaaaacaacc	cttataagca	taagacggac	tacggccatg	ccggcgtgac	35220
cgtaaaaaaa	ctggtcaccc	tgattaaaaa	gcaccaccga	cagctcctcg	gtcatgtccg	35280
gagtcataat	gtaagactcg	gtaaacacat	caggttattt	catcggtcag	tgctaaaaag	35340
cgaccggaaat	agcccgcccc	aatacatacc	cgcaggcgta	gagacaacat	tacagcccc	35400
ataggaggt	taacaaaattt	aataggagag	aaaaacacat	aaacacctga	aaaaccctcc	35460
tgcctaggca	aaatagcacc	ctccgcctcc	agaacaacat	acagcgcttc	acagcggcag	35520
cctaaccatgc	agccttacca	gtaaaaaaa	aaaccttata	aaaaaacacc	actcgacacg	35580
gcaccagctc	aatcagtac	agtgtaaaaa	agggccaaat	gcagagcgag	tatataagg	35640
actaaaaaat	gacgtaacgg	ttaaagtcca	caaaaaacac	ccagaaaacc	gcacgcgaac	35700
ctacgcccag	aaacgaaagc	caaaaaaccc	acaacttctt	caaatcgta	cttcgtttt	35760
cccacggtac	gtaacttccc	attttaagaa	aactacaattt	cccaacacat	acaagttact	35820
ccgcctaaa	acctacgtca	ccggcccggt	tcccacgccc	cgcgccacgt	cacaacttcc	35880
accccctcat	tatcatattt	gcttcaatcc	aaaataaggt	atattatttg	tgatg	35935

<210> 10

<211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> NSsuboptmut

<400> 10

gccaccatgg	cccccatcac	cgcctacagc	cagcagacca	ggggcctgt	gggctgcac	60
atcaccagcc	tgaccggac	cgacaagaac	caggtggagg	gagaggtgca	ggtggtag	120
accgcgtaccc	agagcttcc	ggccacctgc	gtgaaacggcg	tgtgctggac	cgtgtaccac	180
ggagccggaa	gcaagaccct	ggccggaccc	aaggcccta	tcacccagat	gtacaccaat	240
gtggatcagg	atctgggtgg	ctggcaggcc	cctccggag	ccaggagcct	gacacccctgt	300
acctgtggaa	gcagcgaccc	gtacctgg	acacgcccacg	ccgatgtat	ccccgtgagg	360
cgcagggcgc	attctcgccg	aagcctgt	agcccttaggc	ccgtgagcta	cctgaaggc	420
agcagcggag	gaccctgt	gtgtccctt	ggccatgccc	tgggcattt	tgcgcgtgcc	480
gtgtgtacca	ggggcgtggc	caaagccgt	gattttgtgc	ccgtggaaag	catggagacc	540
accatcgca	gccctgttt	caccgacaac	agtcctcccc	ctgcctgtcc	ccaaatcatc	600
cagggtgtc	acctgcacgc	ccctaccg	tctggcaaga	gcaccaaggt	gcccgtgcc	660
tacgcgcgtc	agggtaccaa	ggtgctgt	ctgaaacccca	gcgtggccgc	tacccctggc	720
ttcggcgctt	acatggacaa	ggcccatggc	atcgacccca	acatccgcac	aggcgtgcgc	780
accatccca	cggggatcc	cgtgactac	agcacatcag	gcaagttcc	ggccgatgga	840
ggctgcagcg	gaggagccta	cgacatc	atctgcacg	agtggccacag	caccgacagc	900
accaccatcc	tgggcatttg	caccgtgt	gatcaggcc	aaacagctgg	agccaggctg	960
gtgggtgtgg	ccacagctac	ccctccgtgc	agcgtgaccg	tgcctccatcc	caatatcgag	1020
gagggtggccc	tgagcaacac	aggcgagatc	cccttctacg	gcaaggccat	ccccatcgag	1080
gccatccgcg	gaggcaggca	cctgatctc	tgcacacgca	agaagaagt	cgacgagctg	1140
gctgccaagc	tgagcgact	gggcataa	ccctggcc	actacaggggg	cctggacgtg	1200
tcagtatcc	ccaccatcg	cgatgtgtt	gtggtgcc	ccgacgcct	gatgacaggc	1260
tacacccgg	acttcgacag	cgtgatcgac	tgcacacat	gcgtgaccca	gaccgtggac	1320
ttcaggctgg	acccacactt	caccatcgaa	accaccaccc	tgcctcagga	tgctgtgagc	1380
aggagccaga	ggcgccggac	caccggaaagg	ggcaggccgc	gaatttatacg	ctttgtgacc	1440
cctggcgaaa	ggccctctgg	catgttcgac	agcagcgtgc	tgtgcgagtg	ctacgacgct	1500

ggctgcgctt	ggtacgagct	gacacccgct	gaaaccagcg	tgcgcctgcg	cgcttatctg	1560
aataccctg	gcctgcccgt	gtgtcaggac	cacccggagt	tctgggagag	cgtttcaca	1620
ggactgaccc	acatcgacgc	ccatttcctg	agccagacca	agcaggctgg	cgacaacttc	1680
ccctatctgg	tggcttatca	ggccaccctg	tgtgctaggg	cccaagctcc	accccttca	1740
tgggaccaga	tgtggaagt	cctgatccgc	ctgaagccca	ccctgcacgg	ccctaccct	1800
ctgctgtacc	gcctgggagc	cgtcagaac	gagggtgaccc	tgacccaccc	catcaccaag	1860
tacatcatgg	cctgcatgag	cgctgatctg	gaagtggtga	ccagcacctg	ggtgctggtg	1920
ggaggcgctc	tggccgtct	ggctgcctac	tgcctgacca	ccggaagcgt	ggtgatcggt	1980
ggacgcata	tcctgagcgg	aaggcccgct	atcggtcccg	atcgcgagtt	cctgtaccag	2040
gagttcgacg	agatggagga	gtgtgcccgc	cacccgtcc	acatcgagca	ggcatgcag	2100
ctggccgaac	agttcaagca	gaaggccctg	ggccctgtgc	agacagccac	caaacaggcc	2160
gaagctgccc	ctcccggtgt	ggaaagcaag	tggaggggcc	tggagacctt	ctgggctaag	2220
cacatgtgga	acttcatctc	ttgcattccag	tacccggcc	gactgagacac	cctgcctggc	2280
aaccccgcta	tcggcagcct	atggcccttc	accctgatca	tcacccctcc	cctgaccacc	2340
cagagcaccc	tgtgttcaa	cattctggc	ggatgggtgg	ccgctcagct	ggcccctct	2400
tcagctgttt	ctggcttctgt	gggcgttgc	attggccggag	ccgctgtggg	cagcattggc	2460
ctgggcaaag	tgtgtgtgg	tattctggct	ggctatggcg	ctggcggtgg	cgagccctgt	2520
gtggccttca	aggtgatgag	cgagagatg	cccaagcaccg	aggacctgtt	gaacctgtcg	2580
cctgcattc	tgagccctgg	agccctgggt	gtgggggtgg	tgtgtgtgc	cattctgagg	2640
cgccatgtgg	gaccgggaga	gggcgtgtg	cagtgatga	accgcctgt	cgccttcgccc	2700
tctcgeggaa	accacgttag	cccttaccac	tacgtgcctg	agagcgacgc	cgctgcccagg	2760
gtgaccaga	tcctgagcag	ctgtaccatc	acccagctgc	tgaagcgct	gcaccagtgg	2820
atcaacgagg	actgcagcac	accctgcagc	ggaagctggc	tgagggacgt	gtgggactgg	2880
atctgcaccc	tgtgtaccga	cttcaagacc	tggctgcaga	gcaagctgt	gcccctaactg	2940
cctggcggtc	ccttcttctc	atgccagcgc	ggatacaagg	gcgtgtggag	ggggatggc	3000
atcatgcaga	ccacccgtcc	ctgcccggcc	cataccacag	gcccacgtaa	gaaacggcgc	3060
atgcgcata	tggccctaa	gacccgtcc	aaacacccgc	acggcaccc	ccccatcaac	3120
gcctacacca	ccggaccctg	caccccccgc	cctgtccca	actacagcag	ggccctgtgg	3180
agggtggctg	ccgaggagta	cgtggaggtg	accagggtgg	gagacttca	ctacgtgacc	3240
ggaatgacca	ccgacaaacgt	gaagtgtccc	tgtcaggtgc	ccgcctccga	atttttacc	3300
gaagtggatg	gggtgcgtt	gcacccgtat	gcccctgcct	gttagggccct	gtgcgcgaa	3360
gaagtgaccc	tccaggtggg	ctgtaccatc	tacccgggtgg	gcagccagct	gcctgcgag	3420
cctgagcccg	atgtggccgt	gtgtaccac	atgtgtaccg	accccaagca	catcacagcc	3480
gaaaccgcta	aaaggcgccct	ggccaggggc	tctccctccaa	gcctggcctc	aagcagcgt	3540
agccagctgt	ctgctcccaag	ctgtaccgtt	acccctccac	cccaccacgt	gagcccccgc	3600
cccgacccgt	tcgaggccaa	ctgtgtgtt	ccgcaggaga	tggggccaa	catcacccgc	3660
gtggagagcg	agaacaaggt	gggtgtgtc	gacagcttcg	acccctgcg	cgccgaggag	3720
gacgagcgcg	agggtgacgt	gccccggag	atccctgcga	agagcaagaa	gtttcccgct	3780
gccatggcca	tctgggttag	acccgttac	aaacccctcc	tgtgtggag	ctgtggaggac	3840
cctgattacg	tcgttccagt	gtgtcatggc	tgttctctgc	ctcccataa	agccctctct	3900
atcccacctc	ctaggcgcaa	aaggaccgtg	gtgtgtacag	aaagcagcgt	gagctctgt	3960
ctggccgaac	tggccaccaa	gacccgttgc	agcagcgaga	gctctgcgt	ggacagcggg	4020
acagccaccc	ctctgcctga	ccaggccagc	gacccggccg	ataagggcag	cgatgtggag	4080
agctatagca	gcacccgtcc	cttggaaaggc	gaaaccctggcg	atcccgatct	gagcgatggc	4140
agctggagca	ccgtgagcga	agaggccagc	gaggacgtgg	tgtgtgtcag	catgagctac	4200
acctggacag	gcgtctgtat	cacaccctgc	gctggccagg	agagcaagct	gcccataaac	4260
gccctgagca	acacccgtct	gaggccaccac	aaacatgggt	acgcccaccac	cagcaggctc	4320
gccggactga	ggcagaagaa	gggtgacccct	gaccggcctgc	aggtgctgg	cgaccactac	4380
cgcgatgtc	tgaaggagat	gaaggccaa	gcccggcccg	tgaaggccaa	gctgctgagc	4440
gtggaggagg	cctgcaagct	gacccccc	cacccggcc	agagcaagtt	cggttacggc	4500
gccaaggacg	tgcgcaaccc	gacccggcc	gcccgttgcacc	acatccacag	cgtgtggaa	4560
gacccgttgc	aggacaccgt	gacccccc	gacccacca	tcatggccaa	gaacggagg	4620
ttctgcgtc	agcccgagaa	ggccggccgc	aagcccgctc	gctgtatgt	gtttcccgat	4680
ctggccgtc	gctgtgtgcga	gaagatggcc	ctgtacgac	tggtgagc	cctgcctcag	4740
gtgggtatgg	gctcaagcta	cggttccag	tacagccctg	gccagcgct	ggagttctct	4800

gtgaacacct	ggaagagcaa	gaagaacccc	atgggcttca	gctacgacac	acgctgcttc	4860
gacagcacgg	tgaccgagaa	cgacatccgc	gtggaggaga	gcatctacca	gtgctgcgac	4920
ctggcccctg	aggccaggca	ggccatcaag	agcctgaccg	agcgcctgt	catcgaggc	4980
cctctgacca	acagcaaggg	acagaactgc	ggatacaggc	gctgttagggc	ctctggcgtg	5040
ctgaccacca	gctgtggcaa	caccctgacc	tgctacctga	aggccagcgc	tgccctgtcg	5100
gctgccaagc	tgccaggactg	caccatctg	gtgaacgccc	ctggcctgtt	ggtgatttgc	5160
gaaagcgctg	gcacccagga	agatgtgcc	agccctgcgc	tgttcacccg	ggccatgacc	5220
aggtactctg	ccccctccgg	agacccccc	cagccccaa	acgacacttgc	gctgtatcacc	5280
agctgtcaa	gcaacgtgag	cgtggctcac	gacgcgcgc	gaaagcgctg	gtactacactg	5340
acacgcgatc	ccaccacccc	tctggctcgc	gctgcctggg	aaaccgcctcg	ccatacacccc	5400
gtgaacagct	ggctgggcaa	catcatcatg	tacgccccta	ccctgtggc	tcgcatgatc	5460
ctgatgaccc	acttcttcag	catcctgtcg	gctcaggagc	agctggagaa	ggccctggac	5520
tgcgcgcacg	acggcgttgc	ctacagatc	gagccccctgg	acctgcggcc	aatcatcgag	5580
cgccgcacg	gcctgtctgc	cttcagcctg	cacagctaca	gcccggcga	aattaatcgc	5640
gtggccagct	gtctgcgcaa	actggcgtg	ccttccttcg	gctgtggag	gcatagggtc	5700
aggagcgtga	gggctaggct	gctgaggccag	ggaggcaggg	ccgttacccgt	tggaaagtac	5760
ctgttcaact	gggcccgtgaa	gaccaagctg	aagctgaccc	ctatccctgc	cgctagccag	5820
ctggacacctg	gccccatggtt	cgtggcttgc	tacagcggag	gacacatcta	ccacagcctg	5880
tctcgcgc	gcctcgctg	gttcatgtcg	tgccctgtcg	tgctgagcgt	gggcgtggc	5940
atctacactgc	tgcccaacgg	ctaaaa				5965

<210> 11

<211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric NSsuboptmut

<400> 11

gccaccatgg	cccccatcac	cgcctacagc	cagcagaccc	gcggcctgt	gggctgcac	60
atcaccagcc	tgaccggccg	cgacaagaac	cagggtggagg	gcgaggtgca	ggtgggtgagc	120
accgcacccc	agagcttctc	ggccacatgc	gtaaacggcg	tgtgctggac	cgtgtaccac	180
ggcgccggca	gcaagaccc	ggccggcccc	aaggggccca	tcacccagat	gtacaccaac	240
gtggcgcagg	acctgggtgg	ctggcaggcc	ccccccggc	ccgcgcgc	gaccgcctgc	300
acctgcggca	gcagcgaccc	gtacctgggt	acccgcacg	ccgacgtgat	ccccgtgcgc	360
cgccgcggcg	acagccgcgg	cgcctgtcg	agcccccgc	ccgtgagcta	cctgaagggc	420
agcagcggcg	gccccctgtc	gtgccccagc	ggccacgccc	tgggcattt	ccgcgcggcc	480
gtgtgcaccc	gccccgtggc	caaggccgtg	gacttgcgtc	ccgtggagag	catggagacc	540
accatgcgc	gccccgtgtt	caccgacaa	agcagccccc	ccgcgcgt	ccagagcttc	600
cagggtggccc	acctgcacgc	ccccacccgc	agcggcaaga	gcaccaaggt	gccccccgc	660
tacgcccggcc	agggctacaa	ggtgtctgg	ctgaacccca	gctgtggcc	caccctgggc	720
tccggccct	acatgagcaa	ggcccacggc	atcgcacccca	acatccgcac	cgccgtgcgc	780
accatcacca	ccggcgcggcc	cgtgacccat	agcacctacg	gcaagttcct	ggccgacggc	840
ggctgcagcg	gccccgtcc	cgacateatc	atctgcacg	agtgcacag	caccgacacgc	900
accacccatcc	tgggcacatgg	caccgtgtcg	gaccaggccg	agaccgcgg	cgccccctgc	960
gtgggtgtgg	ccaccgcac	ccccccggc	agcgtgaccc	tgccccaccc	caacatcgag	1020
gagggtggccc	tgagcaacac	cggcgcgc	ccttctacg	gcaaggccat	ccccatcgag	1080
gccatccgcg	gccccgtcc	cctgatctc	tgccacagca	agaagaagt	cgacgagctg	1140
gcccgcacgc	tgagcgccct	gggcacatca	gcccgtggct	actaccgcgg	cctggacgtg	1200
agcgtatcc	ccaccatcg	cgacgtgg	gtgggtggca	ccgacgcctt	gatgacccgc	1260
tacaccggcg	acttcgacag	cgtgtatcg	tgcaacaccc	gctgtgaccc	gaccgtggac	1320
tccggccctgg	accccacctt	caccatcg	accaccaccc	tgccccagga	cgccgtgagc	1380
cgccgcgcgc	gccccggcc	caccggccgc	ggccgcgcgc	gcatctacc	cttcgtgacc	1440
ccccggcgac	gccccagccg	catgttgcac	agcagcgtgc	tgtgcgagtg	ctacgacgcc	1500

ggctgcgcct	ggtacgagct	gaccccccgc	gagaccagcg	tgcgcctgcg	cgcctacctg	1560
aacacccccc	gcctgcccgt	gtgccagac	cacccctgg	tctgggagag	cgtttcacc	1620
ggcctgaccc	acatcgacgc	ccacttctg	agccagacca	agcaggccgg	cgacaacttc	1680
ccctacctgg	tggcttacca	ggccacccgt	tgcgcctgc	cccaggcccc	cccccccccgc	1740
tgggaccaga	tgttggaaatg	cctgatccgc	ctgaagccca	ccctgcacgg	ccccacccccc	1800
ctgctgtacc	gcctgggcgc	cgtcagaac	gaggtgaccc	tgacccaccc	catcaccaag	1860
tacatcatgg	cctgcatgag	cgccgaccc	gaggtggta	ccagcacctg	ggtgttgg	1920
ggccgcgtgc	tggcccccct	ggccgcctac	tgcctgacca	ccggcagcgt	ggtgtatctg	1980
ggccgcatac	tcctgagcgg	ccgcccccc	atcgtggcc	accgcgagtt	cctgttaccag	2040
gagttcgacg	agatggagga	gtgcgcacgc	caccccttgc	acatcgagca	gggcacatgcag	2100
ctggccgagc	agttcaagca	gaaggccctg	ggccctgtgc	agaccgcac	caagcaggcc	2160
gaggccgccc	ccccctgtgt	ggagagcaag	tggcgcgccc	tggagaccc	ctggcccaag	2220
cacatgttgg	acttcatcg	cggcacatcc	tacccctggcc	gcctgagcac	cctggccggc	2280
aaccccgcc	tcggccagct	gatggccctc	accgcacga	tcaccagccc	cctgttaccac	2340
cagagcaccc	tgcgttcaa	catccctggc	ggctgggtgg	ccgcccagct	ggcccccccc	2400
agccgcgc	gccccttcgt	gggcgcggc	atcgcggcg	ccggcgttgg	cagcatcgcc	2460
ctgggcgaagg	tgcgttgg	catccctggc	ggctacggcg	ccggcgttgg	ccggccctg	2520
gtggccttca	aggtgatgag	cggcgagatg	cccacccac	aggaccttgg	gaacctgttg	2580
cccgccatcc	tgagccccgg	cgccctgtgt	gtggcgttgg	tgtgcgcgc	catccctgcgc	2640
cgccacgttgg	gccccggcg	gggcgcgtg	cagtggatga	accgcctgtat	cgcccttcgc	2700
agccgcggca	accacgttag	ccccacccac	tacccctggcc	agagcgcacgc	cgccgcggc	2760
gtgacccaga	tcctgagcag	cctgaccatc	accacgtgc	tgaagccct	gcaccagtgg	2820
atcaacgagg	actgcacgc	ccctgcacgc	ggcagcttgc	tgccgcacgt	gtggacttgg	2880
atctgcaccc	tgcgttccgc	cttcaagacc	tggcttgcaga	gcaagctgtct	gccccagctg	2940
ccggcgttgc	ccttcttcag	ctgccacgc	ggctacaagg	gcgtgttgg	ccggcgcggc	3000
atcatgcaga	ccacccctgc	ctggggcgc	cagatccac	gcccacgttgg	gacccggc	3060
atgcgcac	tggggcccaa	gacccgcacgc	aaccccttgc	acggcacctt	ccccatcaac	3120
gcctacacca	ccggccccctg	caccccccac	ccccccttgc	actacaggcg	cccccctgtgg	3180
cgcgttggcc	ccggaggatg	cgtggaggtg	acccgcgttgg	gcaacttcca	ctacgttgc	3240
ggcatgacca	ccgacaaacgt	gaagtgc	tgcctggatgc	ccgccccccg	gttcccttacc	3300
gagggtggacg	gcgtgcgcct	gcacccgtac	gccccccgc	gcccacgttgg	gacccggc	3360
gagggtgac	tccaggtgg	cctgaaaccag	tacccctgg	gcaagccagct	gcccctgcgag	3420
cccgagcccg	acgtggccgt	gctgaccac	atccgttgc	accccacca	catcaccgc	3480
gagaccgc	agcggccct	ggcccgcggc	agccccccca	gcctggccag	cagcagcgc	3540
agccagctga	gcgcgcgc	cctgaaaggcc	acccgttgc	cccacccac	gagcccccac	3600
ccgcaccc	tcgaggccaa	cctgctgttgc	cgccaggaga	tggcggcaa	catcaccgc	3660
gtggagagcg	agaacaaggt	ggtgttgc	gacagcttgc	accccctgc	ccggcaggag	3720
gacgagcgc	aggtgagct	gcccgcggag	atccgttgc	agagcaagaa	gttcccccgt	3780
gcatgacca	tctgggctag	acccgttac	aaccccttcc	tgtggagag	ctggaggac	3840
cctgattac	tgcctccat	ggtgttgc	tgccttgc	ctcccataa	agcccccttcc	3900
atcccaccc	ctaggcgc	aaggaccgt	gtgttgc	aaagcagct	gagcttgc	3960
ctggccgaa	tggccacca	gaccccttgc	gacccgttgc	gctctgcgt	ggacagcgg	4020
acagccaccc	ctctgcctg	ccaggccac	gaggcaccac	ataaggcc	cgatgttgg	4080
agctatacg	gcacccctcc	cctgaaaggc	gaccccttgc	atcccgatct	gaggatggc	4140
agctggagc	ccgtgacgc	agaggccac	gaggacgttgc	tgtgttgc	catagactac	4200
acctggacag	gcgcctctgt	acacccctgc	gctgcgcagg	agagcaagct	gcccaccaac	4260
gcctgacca	acacccctgt	gaggcaccac	aacatgttgc	acgcccac	cagcaggct	4320
gcccggact	ggcagaagaa	ggtgttgc	gaccgcctgc	aggtgttgc	cgaccactac	4380
cgcgatgtc	tgaaggat	gaaggccaa	gcccaccc	tgaaggccaa	gctgttgc	4440
gtggaggagg	cctgcaagct	gacccccc	cacacgc	agagcaagtt	cggttgc	4500
gccaaggacg	tgcccaac	gacccgttgc	gaccgttgc	acatccac	cgtgttgg	4560
gacccgttgc	aggacaccgt	gaccccccac	gacaccacca	tcatggccaa	gaacggatgt	4620
ttctgcgtgc	agcccgagaa	gggcggccgc	aaggccccc	gcctgtatgt	gttcccccac	4680
ctggccgtgc	gcgtgtgc	gaagatggcc	ctgtacgc	tgtgttgc	cctggcccg	4740
gtgggtatgg	gcacgcac	cggttccac	tacagcccc	gcacgcgcgt	ggagttcc	4800

gtgaacacct	ggaagagcaa	gaagaacccc	atgggcttca	gctacgacac	ccgctgcttc	4860
gacagcaccg	tgaccgagaa	cgacatccgc	gtggaggaga	gcatctacca	gtgctgcac	4920
ctggccccc	aggcccgc	gcccataag	agcctgaccg	agcgcctgt	catggcgcc	4980
ccctgacca	acagcaaggg	ccagaactgc	ggctaccgc	gctgcccgc	cagccgcgt	5040
ctgaccacca	gctgccc	caccctgacc	tgttacactg	aggccagcgc	ccgcctgcgc	5100
gcccaca	tgca	gaggact	caccatgt	gtgaacgc	ccggcctgg	5160
gagagcgc	gcaccc	ggacccc	agcctgc	tgttacacc	ggccatgacc	5220
cgctacagc	ccccccccc	cgccccccc	cagcccg	acgacactg	gctgatcacc	5280
agctgcagc	gcaacgt	cg	tgccccc	gacgc	gcaagcg	5340
accccg	ccaccaccc	cctggcc	gccc	ccatgg	ccacaccc	5400
gtgaacagc	gt	ggctgg	catcat	tacgc	ccctgtgg	5460
ctgatgaccc	actt	ttc	catc	ctgt	ccgc	5520
tgccagatct	acgg	gc	ctac	agc	atcg	5580
cgcc	tc	ct	agc	atc	gatc	5640
gtggccag	gt	ctgc	gat	cc	ccac	5700
cgca	gc	cc	at	cc	ccgc	5760
ctgttca	at	ggcc	gg	cc	ccac	5820
ctggac	ct	gtg	gg	cc	ccat	5880
agccgc	cc	cc	cc	cc	atct	5940
atctac	cc	cc	cc	cc	atct	5965

<210> 12

<211> 10

<212> RNA

<213> Artificial Sequence

<220>

<223> Ribosome binding site

<400> 12

gccaccaugg

10

<210> 13

<211> 49

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic polyadenylation signal

<400> 13

aauaaaaagau cuuuuuuuc auuagaucug uguguugguu uuuugugug

49

<210> 14

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Additional nucleotides present in pVIJns-NS

<400> 14

tcttagagcgt ttaaaccctt aattaagg

28

<210> 15

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Additional nucleotides present in pV1Jns-NSOPTmut

<400> 15
tttaaatgtt taaac

15

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide primer

<400> 16
tcgaatcgat acgcgaacct acgc

24

<210> 17
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide primer

<400> 17
tcgacgtgtc gacttcgaag cgcacaccaa aaacgtc

37

THIS PAGE BLANK (uspto)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/031588 A3

(51) International Patent Classification⁷: C12N 15/40,
15/51, 15/85, 15/86, 15/861, A61K 48/00

[IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT).
COLLOCA, Stefano [IT/IT]; Via Pontina KM. 30.600,
I-00040 Pomezia (IT).

(21) International Application Number: PCT/US02/32512

(74) Common Representative: MERCK & CO., INC.; 126
East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(22) International Filing Date: 10 October 2002 (10.10.2002)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,
VN, YU, ZA, ZM, ZW.

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/328,655 11 October 2001 (11.10.2001) US
60/363,774 13 March 2002 (13.03.2002) US

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. [IT/IT]; VIA PONTINA KM. 30.600, I-00040 POMEZIA (IT).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(72) Inventors; and

(75) Inventors/Applicants (for US only): EMINI, Emilio, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). KASLOW, David, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). BETT, Andrew, J. [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). NICOSIA, Alfredo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LAHM, Armin [DE/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LUZZAGO, Alessandra [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). CORTESE, Riccardo

(88) Date of publication of the international search report:
30 October 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HEPATITIS C VIRUS VACCINE

(57) Abstract: The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

WO 03/031588 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/32512

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/40, 15/51, 15/85, 15/86, 15/861; A61K 48/00
US CL : 514/44; 424/93.2; 435/320.1, 455, 456

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 514/44; 424/93.2; 435/320.1, 455, 456

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,127,116 A (RICE et al.) 03 October 2000 (03.10.2000), column 45, lines 18-57.	1, 2
A	WO 01/30812 A2 (CHIRON CORPORATION) 03 May 2001 (03.05.2001).	1-54
A	WO 97/47358 A1 (MERCK & CO., INC.) 18 December 1997 (18.12.1997).	1-54

<input type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
	Special categories of cited documents:	"P"	Later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 09 July 2003 (09.07.2003)	Date of mailing of the international search report 02 SEP 2003
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Scott D. Priebe Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/32512

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-54

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/32512

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-54, drawn to a nucleic acid encoding a HCV polyprotein.

Group II, claim(s) 55-59, drawn to a chimeric adenovirus vector comprising sequence derived from human adenovirus serotypes 5 and 6.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The technical feature of invention I is a nucleic acid encoding a polyprotein derived from an HCV polyprotein, whereas the technical feature of invention II is a chimeric adenoviral vector comprising a heterologous sequence. These two features are not related. Invention I does not require vector of invention II, nor does is the vector of invention II required to contain the polymucleotides of invention I.

Continuation of B. FIELDS SEARCHED Item 3:

MEDLINE, EMBASE, CAPLUS, BIOSIS, SCISEARCH, USPT, PGPB, DERWENT, GENBANK, GENESSEQ
search terms: HCV, hepatitis C virus, vaccine, NSSB, NSSB near inactiv? or non-functional, SEQ ID NO: 1, SEQ ID NO: 2

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)